ANALYSIS OF REPETITIVE DNA SEQUENCES IN THE MEXICAN AXOLOTL *AMBYSTOMA MEXICANUM*

JOHN ROBERT ADAIR B.Sc (HONS)

SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

UNIVERSITY OF WARWICK

COVENTRY

DEPARTMENT OF BIOLOGICAL SCIENCES

APRIL 1983

CONTENTS

		PAGE
TITLE		
CONTENTS		
LIST OF FIGURES		
LIST OF TABLES		
ACKNOWLEDGEMENTS		
DECLARATION		
SUMMARY		
CHAPTER 1.	INTRODUCTION	1
CHAPTER 2.	MATERIALS AND METHODS	19
CHAPTER 3.	RESULTS AND DISCUSSION I PHYSICAL PARAMETERS OF THE AXOLOTL GENOME	42
CHAPTER 4.	RESULTS AND DISCUSSION II REASSOCIATION KINETIC ANALYSIS OF AXOLOTL REPETITIVE DNA	53
CHAPTER 5.	RESULTS AND DISCUSSION III MOLECULAR CLONING AND ANALYSIS OF AXOLOTL REPETITIVE DNA.	72
CHAPTER 6.	CONCLUDING REMARKS	98
APPENDIX I	PHYLOGENY OF THE AMBYSTOMATIDAE	100
APPENDIX II	REASSOCIATION KINETIC THEORY	103
APPENDIX III	RESTRICTION ENZYME MAPPING OF CLONED DNA	125
REFERENCES		

LIST OF FIGURES

FIG.		FOLLOWING PAGE
2.1	SIZE FRACTIONATION OF SHEARED AXOLOTL DNA	33
2.2	HYDROXYLAPATITE ELUTION PROFILE	36
3.1	ANALYTICAL ULTRACENTRIFUGATION OF AXOLOTL DNA	43
3.2	OPTICAL MELTING CURVE	44
3.3	QUALITATIVE DEMONSTRATION OF METHYLATION AT THE SEQUENCE 5'CCGG	46
3.4	PREPARATIVE CsCl CENTRIFUGATION OF AXOLOTL DNA	50
3.5	G+C DISTRIBUTION OF HIGHLY REPETITIVE AXOLOTL DNA	51
3.6	ANALYTICAL ULTRACENTRIFUGATION OF 2.5 kbp AXOLOTL DNA	51
3.7	HaeIII DIGESTION OF AXOLOTL DNA	52
4.1	REASSOCIATION KINETICS OF SHEARED AXOLOTL E.coli DNA	54
4.2	SCHEMATIC REPRESENTATION OF SEQUENCE INTERSPERSION	60
4.3	HYDROXYLAPATITE BINDING WITH INCREASING TRACER FRAGMENT SIZE	62
4.4	MELTING CURVES OF SHEARED DNA	64
4.5	SIZE DISTRIBUTION OF DUPLEX DNA	66
4.6	2.5 kb REASSOCIATION KINETICS	69
5.1	SCHEMATIC FOR "SHOTGUN" CLONING OF AXOLOTL DNA	73

FIG.		FOLLOWING PAGE
5.2	CHOICE OF RESTRICTION ENZYMES FOR CLONING	74
5.3	SCREENING OF RECOMBINANT CLONES WITH COT 0 - 50 DUPLEX DNA	76
5.4	BamHI DIGESTION OF PUTATIVE AXOLOTL RECOMBINANT PLASMIDS	77
5.5	CURRENT RESTRICTION MAPS OF AXOLOTL RECOMBINANT CLONES	78
5.6	RESTRICTION ENZYME DIGESTS OF SELECTED AXOLOTL CLONES	78
5.7	PLASMID SELF REASSOCIATION	80
5.8	DRIVER-TRACER REASSOCIATIONS	80
5.9	SATURATION HYBRIDISATION	81
5.10	MELTING CURVES OF RECOMBINANT PLASMIDS	82
5.11	ANALYSIS OF DIVERGENCE OF REPEAT FAMILIES	83
5.12	PLASMID CROSS HYBRIDISATION	85
5.13-5.15	GENOMIC HYBRIDISATION OF CLONED DNAs	86
5.16	HYBRIDISATION OF CLASS I REPEATS TO AXOLOTL LAMPBRUSH CHROMOSOME	94
5.17	HYBRIDISATION OF pAMW1184 TO AXOLOTL LAMPBRUSH CHROMOSOMES	94
5.18	PARTIAL KARYOTYPES OF LAMPBRUSH CHROMOSOME PREPARATIONS HYBRIDISED WITH ^{3}H pAMW1184	94
AI.1	GEOGRAPHICAL DISTRIBUTION OF AMBYSTOMA SPECIES	102
AII.1	RELATIVE REASSOCIATION RATE : DEPENDENCE UPON %G+C	117
AIII.1	RESTRICTION MAPPING : CLONE 1184	125
AIII.2	AID TO RESTRICTION MAPPING : pAT153 FRAGMENT SIZES	126

AIII.3 PARTIAL MAPPING DATA 127

LIST OF TABLES

TABLE		FOLLOWING PAGE
1.1	INTERSPERSION PATTERNS IN INSECTS	2
3.1	C VALUE DETERMINATIONS	42
3.2	%G+C CONTENT IN ANIMALS	48
4.1	KINETIC DATA	55
4.2	URODELE REASSOCIATION KINETIC DATA	57
5.1	RESPONSE OF CLONES TO GENOMIC PROBES	77
5.2	PLASMID SELF REASSOCIATION	80
5.3	DRIVER-TRACER REASSOCIATIONS	80
5.4	ESTIMATE OF REITERATION FREQUENCY BY SATURATION HYBRIDISATION	81
5.5	THERMAL ELUTION DATA	82
5.6	DIVERGENCE DATA	83
AI.1	CLASSIFICATION OF THE URODELES	101
AI.2	CLASSIFICATION OF THE AMBYSTOMATIDAE	101

ACKNOWLEDGEMENTS

I wish to acknowledge the assistance of the following in the preparation of this thesis : my supervisor Dr. Marcela Vlad, for constant guidance and criticism, for help with the cytological experiments, and for karyotype analysis of several lampbrush chromosome preparations; Professor H. R. Woodland for microinjection of axolotl embryos; Dr. R. J. Avery for instruction in the use of the Gilford spectrophotometer and thermoprogrammer; Dr. C. S. Dow for instruction in the use of the Model E. Analytical Centrifuge; Professor H. C. MacGregor and Mrs. H. Horner for instruction in the use of the Vickers Integrating microdensitometer at Leicester University; Mr. N. Coveney of the U.O.W.C.U. for assistance in adapting computer programs for use of the University Burroughs B6700 computer; the members of the Developmental Biology group, in particular Dr. Vaughan Hilder, Gail Dawson and Nigel Livesey of the 'newt group', for suggestions and criticism; Christine Newton and Elizabeth Ballantine for the smooth running of the lab; Mrs. West for her excellent typing; and the axolotls without whom none of it would have been possible.

I also wish to thank my colleagues at G. D. Searle & Co. Ltd., Dr. A. G. Porter and Dr. P. G. Boseley, for their patience during the completion of the manuscript.

Finally, I wish to thank my fiancee, Yvonne, for her encouragement and cheerfulness despite it all.

DECLARATION

I declare that the work described in this thesis is my own, except where indicated in the text and in the acknowledgements already made. No part of this thesis has been presented for another degree.

J. ADAIR

SUMMARY

The mexican axolotl Ambystoma mexicanum is a member of a family of salamanders with large nuclear genomes. The known geographical distribution and evolutionary history of the genus Ambystoma makes it an ideal model system for the study of the evolution of repetitive DNA sequences in large genome animals.

The genome of the axolotl has been examined by various methods, in particular by reassociation kinetic analysis and associated techniques on bulk DNA and by the molecular cloning of several members of the repetitive fraction of the genome. The haploid genome (37 pg) can be resolved into several distinct components. A foldback fraction (4.6%), a moderately repetitive fraction (43.3%), a slow repetitive fraction (20.4%) and a unique fraction (31.7%). Foldback regions appear to form randomly dispersed clusters. Members of the moderately repetitive fraction are predominantly interspersed amongst other members of the same fraction. At moderate fragment length only (57.5%) of the genome is arranged as moderately repetitive DNA interspersed with more slowly repeating sequences.

Fragments of axolotl DNA have been cloned by recombinant DNA techniques. Clones containing axolotl DNA were selected. One clone (pAMW1131) has an apparent repetition frequency of 1480 in the axolotl genome and may be part of a larger fragment undergoing reamplification in the genome. A second clone (pAMW1184) is highly repeated with a short repeat length. Some family members appear to be scattered through the genome. Some members are transcribed on lampbrush chromosomes. A third clone (pAMW1199) has several features characteristic of bulk repetitive DNA. It has a complex organisation in the genome and may contain a number of shorter repeats present in the genome at different frequencies. These clones represent different aspects of the organisation of the axolotl genome. Together with the general genome analysis they provide a starting point for comparative evolutionary analyses of the genus Ambystoma.

ABBREVIATIONS

EDTA	Ethylenediaminetetraacetate
POPOP	1,4-bis(5-phenyloxazolyl-2)-benzene
PPO	2,5-diphenyloxazole
TES	3-[tris(hydroxymethyl)methyl]amino-propanesulfonic acid
TRIS	tris(hydroxymethyl)aminomethane

ABBREVIATIONS

Standard abbreviations for semi-systematic or trivial names have been used in this thesis. For reference see 'Instructions to Authors' Eur. J. Biochem. Vol. 131 (1) 1983.

"It had begun with a leaf caught in the wind, and it became a tree; and the tree grew, sending out innumerable branches, and thrusting out the most fantastic roots. Strange birds came and settled on the twigs and had to be attended to. Then all around the tree, and behind it, through the gaps in the leaves and boughs, a country began to open out; and there were glimpses of a forest marching over the land, and of mountains tipped with snow".

From Tree and Leaf by J. R. R. Tolkien

CHAPTER 1

INTRODUCTION

Since the mid 1960's work by many groups has led to the conclusion that a portion of the genome of most eukaryotes consists of sequences with a wide range of repetition. These include the satellite DNAs, which may have millions of copies of a basic sequence per genome and which, due to their tandem arrangement, may be isolated easily from the rest of the genome by isopycnic centrifugation or restriction enzyme cleavage (reviewed in Lewin, 1981; Peacock et al, 1980; Maio et al, 1977; Dover, 1977). At the other end of the spectrum repeat families with only a few (tens to hundreds) of copies can be identified (see for example, Bonner et al, 1973; Graham et al, 1974; Moore et al, 1978, 1980; Anderson et al, 1981). Between these extremes there is a complete spectrum of repetition frequency.

The formation, evolution and possible function of these sequences and their organisation with respect to the coding fraction of the genome has been the subject of extensive investigation in the past two decades. The result has been the accumulation of a vast amount of data on genome organisation in eukaryotes which is constantly being supplemented.

This introduction will attempt to describe those processes which may be involved in the amplification, dissemination and evolution of eukaryotic repeated sequences. This approach will, hopefully, lead to an examination of how these processes may be involved in the evolution of genome size.

Much of the early work on the organisation of repeated sequences was based on a desire to understand gene organisation and regulation in complex systems. Repeated sequences offered the means by which dispersed gene sequences could be coordinately regulated in a controlled manner through

the development of an organism (see Britten and Davidson, 1969). Initial studies on the detailed sequence organisation of Xenopus laevis indicated that short repeated sequences of around 300 bp were interspersed amongst single copy sequences of around 1,000 bp (Davidson et al, 1973). This corresponded with then current ideas on gene lengths based on mRNA sizes and so generated intense study in other systems. It was quickly shown that similar organisations existed in sea urchins (Graham et al, 1974) and rats (Bonner et al, 1973) suggesting a widespread occurrence of this type of organisation. However, a different pattern of organisation was claimed for the insect Drosophila melanogaster (Manning et al, 1975). In this organism the interspersed repeats had an average length of 5.6 kbp with a range from 0.5 to at least 13 kbp. The single copy DNA adjacent to these repeats averaged more than 13 kbp. Although it is possible that the procedures adopted to analyse the genome may have selected against short interspersed repeats and single copy DNA (Wensink, 1977; Wensink et al, 1979) a pattern qualitatively similar to that seen in Drosophila was soon discovered in several other insect species including both dipterans and hymenopterans (see Table 1.1). However some insect species were also shown to have an organisation similar to that of Xenopus (see Table 1.1). This apparent disparity in organisation between closely related species was of great concern, particularly as it had been shown that there was evidence, at least in sea urchins, that coding sequences were adjacent to short repeats (Davidson et al, 1974).

As more data emerged a picture of genome organisation throughout the eukaryotes began to develop. In the most primitive eukaryotes few repeated sequences could be distinguished. Yeasts and fungi typically consist of 80 - 98% unique sequences. The majority of the repeats could be accounted for as coding for rRNA (e.g. Whitney and Hall, 1974; Hudspeth et al, 1977; Ullrich et al, 1980; Lauer et al, 1977; Timberlake, 1978; Dons et al, 1979; Dons and Wessels, 1980; Krumlauf and Marzluff, 1979). No evidence could

TABLE 1.1

INTERSPERSION PATTERNS IN INSECTS

ORGANISM	C VALUE (pg)	REFERENCE	
a) LONG PERIOD			
DROSOPHILA MELANOGASTER	0.15	MANNING *et al*	(1975)
		CRAIN *et al*	(1976a)
CHIRONOMUS TENTANS	0.15	WELLS *et al*	(1976)
APIS MELLIFERA	0.35	CRAIN *et al*	(1976b)
SARCOPHAGA BULLATA	0.61	SAMOLS & SWIFT	(1979)
b) SHORT PERIOD			
MUSCA DOMESTICA	0.89	CRAIN *et al*	(1976b)
ANTHERAEA PERNYI	1.0	EFSTRATIADIS *et al*	(1976)

be determined, at least by the techniques used, for interspersion of unique sequences with the repeat elements. In Achyla, for example, the unique sequences were calculated to be at least 1.35×10^5 bp long (Hudspeth et al, 1977). Similar organisation was noted in Chlorella pyrenoidosa (Bayer and Dalmon, 1975) and Chlamydomonas reinhardtii (Howell and Walker, 1976).

However in the dinoflagellate, Crypthecodinium cohnii, a short period - Xenopus type pattern was observed (Allen et al, 1975; Hinnebusch et al, 1980). Euglena gracilis was shown to have short period pattern (Rawson et al, 1977) as was the cellular slime mould Dictyostelium discoideum (Fritel and Kindle, 1975), Amphioxus (Schmidtke et al, 1979), and all echinoderms so far studied (e.g. Graham et al, 1974; Brykov et al, 1978). It was suggested that interspersion of repeated sequences developed as a (Xenopus type) short period pattern early in eukaryotic evolution. Plant genomes in general show short period interspersion where these have been analysed, e.g. Walbot and Dure, 1976; Zimmerman and Goldberg, 1979; Murray et al, 1978; Flavell et al, 1974, 1979; Flavell and Smith, 1976.

Apart from the insects noted the only other group to exhibit extensive Drosophila-like long period interspersion are birds (Arthur and Strauss, 1978; Eden and Hendrick, 1978; Epplen et al, 1978, 1979), while the closely related reptiles show short period interspersion (Epplen et al, 1979), as do all amphibia and fish so far studied (e.g. Davidson et al, 1973; Sommerville and Malcolm, 1976; Baldari and Amaldi, 1976, 1977; Bozzoni and Beccari, 1978; Graham and Schanke, 1980; Schmidtke et al, 1979; Serra and Mandarino, 1979; Hanham and Smith, 1979).

When the genomic interspersion patterns of the birds and insects were being examined it was suggested (Crain et al, 1975) that the Drosophila pattern of interspersion may have developed from a more general Xenopus-like pattern by loss of those sequences involved in short period interspersion, due to selection for a reduction in genome size. This suggestion

allowed the possibility of a distinction between repeated sequences intimately connected with gene control and expression and those, in most cases the majority, which play no obvious part in gene expression.

This second class of repeats is increasingly receiving more study as it is becoming clear that the eukaryotic genome is a dynamic rather than a static structure. It is becoming apparent that a small number of underlying molecular events may be responsible for the wide range of observations on repetitive sequence organisation. Of these perhaps the most important are those which allow the multiplication and dissemination of sequences throughout the genome and those which affect the subsequent evolution of repeat family members.

REPEATED SEQUENCES : FORMATION, DISSEMINATION, EVOLUTION

(a) Formation

The formation of repeat sequences required a mechanism which will multiply a particular sequence rapidly, in germ line tissue. The amplified sequences must persist in the genome either by acquiring a function or by existing "selfishly" (see Orgel and Crick, 1980; Doolittle and Sapienza, 1980). Several models for sequence amplification exist. These are not mutually exclusive. They include :

1) Rolling circle mechanisms
2) Unequal crossing over events
3) Saltatory replications
4) Replicative loops
5) Short range replication errors
6) Transposition mediated duplications.

Each of these processes is capable of producing duplicate copies of sequence. The rolling circle is the mechanism proposed to explain the large scale amplification of ribosomal DNA sequences which occurs during

amphibian oogenesis (Hourcade et al, 1973). Similarly in methotrexate treated tissue culture cells the amplification of dihydrofolate reductase genes is believed to be achieved by a rolling circle intermediate (Alt et al, 1978). In this case reintegration back into the chromosome of amplified sequences occurs. This mechanism allows the possibility of large scale amplification over a short time period.

Unequal crossing over between sister chromatids may be a major factor in the development of tandemly arranged sequences (see Smith, 1973, 1976). Unequal crossing over undoubtedly occurs in eukaryotic cells and there is genetic and cytological evidence in support of the mechanism. There is however little evidence for cross overs within heterochromatin, the most common location for satellite DNA - tandemly repeated highly repetitive DNA.

The observations of Keyl (1965 a + b) that entire chromomeres may be selectively over replicated in Chironomus leads to the possibility of amplification at the level of the chromatin domain. In an organism such as Triturus, where the unit of replication may be ten times that of Xenopus, (Callan, 1972) duplications of 5×10^5 base pairs may occur.

On a smaller scale multiple reinitiations of replication, which are suggested by the observations of Keyl (1965) and hypothesised by Britten and Kohne (1968), would lead to saltatory (disproportionate) amplification of sequences around replication origins. Similarly errors in replication or replication repair mechanisms are capable of producing adjacent short tandem repeats.

Recently it has been suggested that sequence amplification may occur by a transposition mechanism analogous to that found in prokaryotes (see Shapiro, 1979; Calos and Miller, 1980). By this mechanism a copy of a duplicated sequence is inserted at a new genomic site. Spradling and Rubin (1981) suggest that such a mechanism might explain the dispersed, apparently conserved, long repeats in Drosophila although other models,

for maintaining sequence homogeneity in dispersed sequences, have recently been described (see later).

Transposition and concomitant sequence duplication via an RNA intermediate has recently been proposed to explain the appearance of "processed genes" (Hollis et al, 1982) in various mammals (Lomedico et al, 1982; Nishioka et al, 1980; Vanin et al, 1980; Hollis et al, 1982) and the newt Notopthalmus viridescens (Stephenson, 1981; cited in Van Arsdel et al, 1981). Similarly, the structure of some sequences coding for small nuclear RNAs and some members of the major human repeat family, the "Alu" family, suggests transposition via an RNA intermediate (reviewed in Van Arsdell et al, 1981). Each of these events would serve to increase the copy number of a sequence family.

Finally, DNA sequences may be amplified by aneuploidy and polyploidy. These processes allow the duplication of whole chromosomes or complete genomes. There are many examples, particularly amongst plant species, of these occurrences (see for example Rees, 1972; Rees and Jones, 1977).

(b) Dissemination

When sequences are amplified they may remain together in a tandem arrangement or be scattered throughout the genome. The mode of amplification may influence the dissemination of sequences, for example amplification via transposition mechanisms automatically indicates sequence dispersal whereas saltatory replication or unequal crossing over events allow the development of large tandem arrays in situ. That sequences do become relocated through the genome after amplification is now well accepted. Recently the mechanisms have begun to be analysed and in some cases data has been put forward as visible evidence of sequence dispersal in operation.

From work in yeast and Drosophila it has become apparent that there is a class of repeat families which share structural similarity to the

integrated proviruses of retroviruses (reviewed by Spradling and Rubin, 1981). Probably the most famous of these is the "copia" family (Finnegan et al, 1978; Spradling and Rubin, 1981). In Drosophila such sequences have been shown to be polymorphic in number and location between different stocks of D. melanogaster, between melanogaster and sibling species, and in cell line stocks (Spradling and Rubin, 1981). Similarly in yeast, the TY-1 element has been shown to be mobile during the propagation of a single S. cerevisiae clone (Cameron et al, 1979).

Mobilisation of these sequences has features common to prokaryotic transposons and to retroviruses in that integration is accompanied by the duplication of a few base pairs present once only in the target site and found immediately adjacent to the ends of the element after insertion. These duplications appear to be a consequence of the integration process (see Galas and Chandler, 1981).

Recently copia elements have been identified in closed circular DNA form, (Flavell and Ish-Horowicz, 1981). These circles may be functionally analogous to the circles of retroviral DNA believed to be intermediates in proviral integration. Circular DNAs in Drosophila have also been shown to contain moderately repetitive DNA sequences other than copia (Stanfield and Lengyel, 1979). Small circular DNAs were first described in Hela cells (Smith and Vinograd, 1972). Members of the predominant human repeat family, the Alu family, have also been shown to be present on circular DNA (Calabretta et al, 1982). Circular DNAs have also been reported to occur in chickens (de Lap and Rush, 1978) where there is evidence for the presence of circularly permuted repeat arrays throughout the genome which have characteristics of having had a circular intermediate (Musti et al, 1981).

It is thus possible that repeat arrays may be excised, circularised, duplicated and then reintegrated at a separate location. Precedent for each step in the process clearly exists.

The integration does not necessarily require a mechanism analogous to that of transposons. It has been recently shown for SV40 that integration into the rat chromosome involves an homologous recombination between a five base pair sequence shared by the two genomes (Stringer, 1982). Thus the sequence homology required for homologous recombination is quite small. Lania et al, (1982) speculate that homologous recombination between relatively closely linked members of a repeat family may allow excision and reintegration at other points in the genome, either by subsequent homologous recombination or by transposition like integration.

It has been demonstrated that in the gene conversion event which occurs during the switching of the mating type genes in Saccharomyces cerevisiae homologous recombination takes place between sequences located at the ends of the mating type genes, without the requirement for internal sequence homology. It has been shown that plasmid sequences inserted between the two homologous regions can also be efficiently transposed. The process can also be reversed by conversion in the opposite direction leading to splicing out of the inserted sequence (Haber and Rogers, 1982). By this mechanism it is possible to consider the duplication - transposition of any sequence bounded by two direct repeats to a location previously occupied by one copy of the repeat, and vice versa. The transposition of Tyl, via the terminal δ sequences, to other locations in the yeast genome could occur by this means (see Cameron et al, 1979).

A complete spectrum of mobilisation appears possible in the mechanisms briefly outlined above. Taken together with the previously described mechanisms which exist for sequence amplification it is immediately clear that virtually any combination of amplification and dispersal can, in theory, occur. It is therefore not difficult to understand situations in which rapid copy number changes can be seen to have occurred between related sequence families in closely related

species, for example, the expansion of satellite DNA sequences in sibling species of Drosophila (Peacock et al, 1981) and other species (e.g. Fry and Salser, 1977; reviewed in Dover, 1977).

(c) Evolution

One further feature of eukaryotic genome organisation which is central to the discussion is that of sequence divergence, that is, the extent to which members of a particular repetitive sequence family have acquired mutations which cause them to be different from other members of the same family. These mutations may be single base changes (i.e. substitutions, insertions or deletions) or may involve short nucleotide stretches (see for example Posakony et al, 1981). The degree of divergence can be determined from the thermal stability of reassociated family members. Thus the conservation or otherwise of a sequence family may be directly ascertained.

It has long been understood that definitions of repetitive and single copy are functional in that for the most part they depend upon the criterion of the method of measurement. Thus in standard reassociation experiments using 60°C and 0.12M sodium phosphate a sequence with 50% G+C may be up to 30% different from a related sequence, yet they will still reassociate. In the same way sequences of up to 70% similarity would not be capable of reassociating under these same conditions. Therefore by varying the reassociation conditions the apparent sequence organisation of a genome can be drastically changed (see for example Murray et al, 1978; Graham and Schanke, 1980). For a given criterion the effect of mutation would be to reduce the apparent copy number of a family and to increase the single copy fraction of the genome. It can be suggested therefore that the oldest families would be the most diverged. However this suggestion does not necessarily hold, for several reasons. Firstly, during the course of evolution a repeat sequence family, or part

of a family, may acquire a function dependent upon its sequence. Thus selection would operate to maintain the sequence of the family or sub-family. There are many examples in which non-diverged sequences can be shown to be common to species separated by long time periods from a common ancestor (e.g. Mizuno et al, 1976; Harpold and Craig, 1977; Moore et al, 1978, 1980; Flavell et al, 1980; Flavell, 1980). Secondly, the processes of sequence amplification noted earlier allow the possibility of reamplification, at any point in the evolutionary history of a family, of family members, to form sub-families (reviewed by Flavell, 1980). This would tend to reduce the apparent age of the family. The reamplification "history" of certain families, particularly satellite DNAs where the situation is magnified, can be traced by mapping emergence or disappearance of restriction sites within sub-families (see for example Peacock et al, 1980). Thirdly, it has recently been suggested that sequence families, whether tandemly arranged or dispersed, may be subject to a form of concerted evolution in that a process analogous to gene conversion may be capable of maintaining sequence homogeneity, without change in family copy number (Brown and Dover, 1981; Coen et al, 1982). In those cases where the situation has been examined it would appear that within a repeat family variant representatives, which may be recognised by particular restriction enzyme motifs, can spread through the population leading to a situation in which distinct polymorphisms can be seen. In related species different sub-family types dominate the family indicating the speed of the convergence process (Brown and Dover, 1981).

The consequences of these processes for gross genome organisation are wide ranging. For example, it has been shown that repeat sequences may be interspersed amongst each other (Wensink et al, 1979; Musti et al, 1981; Scheller et al, 1981) and that these arrays may appear as single copy DNA depending upon the criterion of reassociation (Murray et al, 1978; Graham and Schanke, 1980). It is not difficult to envisage a

situation in which a repeat element within an array which is apparently single copy can become converted to a more predominant variant of the same family. In this instance a stretch of apparently single copy DNA would be replaced by a sequence showing repeat interspersion as measured by reassociation kinetics and classical interspersion analysis (see Davidson et al, 1973, and below).

Given that the processes of multiplication, dissemination and sequence evolution described above are general phenomena, it is possible to re-examine eukaryotic sequence organisation to try and discern, within these trends, the way in which the organisation of a particular genome has been shaped during its evolutionary history. One fundamental facet of genome evolution which is particularly bound up with the processes described above is that of genome size.

Evolution of genome size

The size of the genome for a particular eukaryote has long posed a great problem in the understanding of genome organisation, function and evolution. The term 'C value' was first used by Swift (1950) to denote the 'constant' haploid amount of DNA for any particular organism. Genome size has been determined by various methods (see Bennett and Smith, 1976 for review) for a large number of organisms. From these data several observations can be summarised as follows :

1. There is a large variation in the haploid DNA content of eukaryotes, ranging from 5×10^{-3} picograms (pg) in some yeasts to 200 pg in certain dinoflagellates (reviewed in Cavalier-Smith, 1978). The smallest eukaryotic values (e.g. yeasts, 0.005 pg) are lower than the largest prokaryotic genome sizes (e.g. cyanobacteria, 0.05 pg) (Cavalier-Smith, 1978).

2. Between groups of organisms the minimum C value for a particular group increases with increasing complexity (Sparrow et al, 1972). An

alternative way of expressing this observation is that there appears to be a minimum value for each group below which the C value cannot be reduced while still retaining the complexity inherent in the group. However the maximum value for a group varies enormously, e.g. algae (0.04 - 200 pg), protozoa (0.05 - 250 pg), amphibia (15 - 100 pg). It is this lack of correlation between developmental complexity and C value that has become known as the C value paradox. (Thomas, 1971; Cavalier-Smith, 1978).

3. Within a taxonomic group the observed C values are not distributed equally across the range. Rather, a modal value exists and the observed values define a smooth frequency distribution around the mode (Bachmann et al, 1972). In many cases the mode is towards the lower end of the distribution (Hinegardner, 1976). Often the more specialised members, and those considered to be more advanced within a group have lower C values than the generalised forms (Bachmann et al, 1972; Morescalchi, 1980).

4. DNA content of a cell has been shown to be correlated with various cellular characteristics, e.g. nuclear volume, cell volume, cell surface area, developmental rate (reviewed by Bachmann et al, 1972; Hinegardner, 1976; Rees and Jones, 1977, 1982; Cavalier-Smith, 1978; Morescalchi, 1980; Bennett, 1982).

5. By comparison between closely related species it can be shown that changes in C value are predominantly due to changes in the relative amounts of the moderately repetitive fraction of the genome (Flavell et al, 1974; Rees and Jones, 1977; Rees and Narayan, 1977; Hutchinson et al, 1980; Straus, 1971; Mizuno and MacGregor, 1974; Baldari and Amaldi, 1976; Morescalchi, 1980).

Thus organisms with the lowest C values within a group generally have the lowest proportion of repeated sequences in the genome. In many cases an increase in the 'single copy' fraction of the genome is also

correlated with increased genome size (see Hutchinson et al, 1980, for review). However as the boundary between 'repetitive' and 'single copy' sequences is criterion sensitive this increase is probably a reflection of the fraction of diverged repeats rather than de novo formation of unique sequence DNA.

From these observations it would appear that there are some constraints upon the C value, that the C value can be correlated to several cellular parameters, and that the moderately repetitive fraction of the genome provides the required evolutionary plasticity.

The way in which the C value of a species evolves has been the object of intense debate recently, particularly since the term "selfish DNA" became fashionable (e.g. Orgel and Crick, 1980; Doolittle and Sapienza, 1980). It has been argued by some that at some point in the life cycle of an organism a parameter correlated to the C value becomes limiting and that in these circumstances alteration in the C value between individuals can be selected for, either for an increase or decrease in the C value (Cavalier-Smith, 1978, 1980; Bennett, 1982). Others have argued that the only selection pressure that DNA experiences directly is the pressure to survive within cells. Therefore if there are ways in which mutation can increase the probability of survival of a sequence within cells without effect on phenotype, then sequences whose only function is self-preservation will arise and be maintained by what has become known as non-phenotypic selection (Doolittle and Sapienza, 1980). Transposable elements and the yeast 2 μm circle have been indicated as possible examples of selfish replicators. In this scheme there is indeed little advantage in selecting against selfish replicators while they constitute a small proportion on the genome, especially if members of a family become dispersed through the genome. However a large proportion of the genome of most eukaryotes is made up of sequences of no apparent function and if this was all selfish DNA then it would appear likely that there would be a significant

selective disadvantage. It has therefore been proposed that the net effect will be that the load of "selfish DNA" will be maintained at a level which can be accommodated without producing a large selective disadvantage. This level will depend on the lifestyle of the organism particularly on those organisms for which DNA replication comprises a large fraction of the total metabolic energy expenditure (Orgel and Crick, 1980; Doolittle and Sapienza, 1980; Doolittle, 1982).

The main difference between these two views is that one suggests that positive selection for increases in C value may occur in certain cases while the other suggests that the natural replicative nature of DNA means that DNA increase is the inevitable result, thus no selection for increase is required. Both views accept that selection for decreases in C value are likely to occur.

The mechanisms by which sequence copy numbers can increase have already been described. However, in many cases the same mechanisms allow for a reduction in copy number. For example unequal crossing over events in tandem arrays result in one amplified and one diminished array. The smaller array may be selected for. Similarly it has been shown that non-homologous gene conversion at the yeast mating type locus can result in both introduction or excision of sequences (Haber and Rogers, 1982). It has already been suggested how this mechanism may allow the amplification of sequences between two directly repeated sequences. The mechanism can quite feasibly work in the opposite direction, as has indeed been shown in yeast. Excision events involving homologous recombination between spaced direct repeats to form a circular DNA containing one copy of the repeat plus the intervening sequences can be directed towards sequence loss by segregation. The reduction in copy number of extrachromosomal dehydrofolate reductase genes on removal of selective pressure (Schimke, 1980) may be a useful analogy. Finally the loss on introns from coding sequences by reintegration of DNA copies of partially or wholly processed RNA transcripts (and subsequent loss of the processing machinery) may

represent an extreme of streamlining of the genome which has been adopted by prokaryotes (Doolittle and Sapienza, 1980).

The consequence of these events of amplification, transposition, divergence, homogenisation and deletion leads to the apparent confusion in the ideas about genome organisation patterns described earlier. However it is not difficult to visualise how these mechanisms may act to produce the patterns described. Data on the fine organisation of low C value organisms has provided much of the evidence for the fluidity of the genome (e.g. yeasts, Drosophila species, sea urchins), however little detailed information is available for high C value organisms, exceptions being the extensive work in Graminae sequence organisation (Smith and Flavell, 1974; Flavell et al, 1979, 1980; Flavell and Smith, 1976; Rumpau et al, 1978, 1980) and work on various Urodele species (see below). Of those high C value systems available for study the Urodeles perhaps offer the most scope.

The range of C value exhibited by the Urodeles is wide (15 - 100 pg; Sommerville, 1977; Cavalier-Smith, 1978; Morescalchi, 1980) and in many cases there already exists extensive phylogenetic and cytogenetic data upon which to base comparative studies of genome evolution.

In perhaps the most detailed comparative study, of the Plethodontid salamanders, (Mizuno and MacGregor, 1974; Mizuno et al, 1976; MacGregor et al, 1976) it was shown that the fraction of repeated sequences was correlated with the C value of each species. Further, when the repetitive fraction of one particular species was used to analyse other species it was found that there was a correlation between the phylogenetic relatedness and the fraction of the repeated sequences common to both species. The repetitive sequences common to any particular pair of species probably represent overlapping subsets of the repetitive component for any one species. For example, P. cinereus and P. dunni have around 10% of their repeated sequences in common (Mizuno et al, 1976). A similar

value can be obtained for P. cinereus - P. vandykei. However, only 75% of the repeats common to P. cinereus - P. dunni are found in P. vandykei. Given the C value differences and different reassociation profiles it can be concluded that each species appears to be in the process of developing a different repetitive sequence profile, this being more noticeable in those species with the higher C value (MacGregor et al, 1976). Between families of genera there appears to be very little conservation of repeat sequence. However the seasonal availability and small number of oocytes per ovary in mature animals for lampbrush chromosome studies, and the difficulty of laboratory maintenance, manipulation and mating makes them a difficult and expensive experimental system.

The Ambystomatid salamanders offer a useful alternative system for study. These salamanders are characterised by a high C value (Morescalchi, 1980). A detailed history of the family exists from which species relationships can be easily deduced (Tihen, 1958; see Appendix I).

There are twenty-six living species, of which twelve are found in Mexico, the rest being distributed throughout the North American continent. The species are divided into three subgenera (see Appendix I). The species can be distinguished by many criteria, of which probably the most useful for evolutionary studies are the species distributions and those biological factors which may affect interbreeding, e.g. habitat, time of mating, mating behaviour, neoteny. In these respects it can be expected that many of the species of the genus would be incapable of natural cross-breeding whereas in others this may be possible (and has been demonstrated to occur, e.g. the A. jeffersonianum complex, and A. mexicanum - A. tigrinum laboratory matings). Thus the genus provides a means of analysing both recent and ancient evolutionary events.

Of the Ambystomatids the mexican axolotl A. mexicanum appears to be an obvious choice to begin comparative analyses. The axolotl has a restricted habitat around Lake Xochimilcho in Mexico so that the effective

breeding pool has been small and inbred. Subsequent laboratory matings in the planned breeding programme at Indiana University and elsewhere have led to the development of distinct strains of axolotl which may be useful in deducing rapid evolutionary changes (see for example Sinclair *et al*, 1977) and which may have helped to further reduce heterozygosity at chromosomal locations which would aid *in situ* hybridisation experiments.

The karyotype of the axolotl has been well characterised from mitotic and lampbrush chromosomes (Callan, 1966). The karyology of certain other Ambystomatids has also been examined (Kezer *et al*, 1980; Sessions, 1982; reviewed in Morescalchi, 1980) and several common features can be determined. It is also possible to form interspecific hybrids between species.

In addition to these immediately useful traits it has been shown that axolotl oocytes can be used for micro-injection studies (Brothers, 1976; Briggs and Cassens, 1966). Further, a large number of mutant types have been discovered including mutations affecting oogenesis, early development, organogenesis, cell and tissue function and adult phenotypic characters. Finally the axolotl exhibits extensive neoteny. The observed correlation between paedogenesis and genome size (Morescalchi, 1980) could be studied using the axolotl as a model.

The axolotl can serve as a basis not only for the study of the relationship of repeated sequences to the C value paradox, but also for the study of the evolution of the karyotype, the relationship of neoteny to genome size and a wide range of developmental processes in addition to the more well known use as a model system in regeneration processes and pattern formation.

AIMS OF THE PROJECT

The broad aims of the project are to analyse in detail the organisation of the genome of the mexican axolotl *Ambystoma mexicanum*,

in particular the repetitive sequence fraction of the genome, by means of reassociation kinetics, molecular cloning, and *in situ* hybridisation studies.

It is proposed to isolate individual sequences containing axolotl repetitive DNA by the techniques of molecular cloning, and to analyse these sequences by a variety of techniques including restriction enzyme mapping, Southern blot hybridisation to genomic DNA, reassociation kinetics and *in situ* hybridisation to mitotic and lampbrush chromosome preparations.

By such analyses it is hoped that a detailed insight into the organisation of the genome may be obtained, which together with the availability of pure sequence probes, will provide a starting point for intra and inter species comparative studies on genome evolution.

CHAPTER 2

MATERIALS AND METHODS

ANIMALS

Mexican axolotls (Ambystoma mexicanum) were obtained either from Gerrard and Haig Ltd., or were a gift from Dr. J. Slack of the Imperial Cancer Research Fund, Mill Hill, London.

Animals were kept in dechlorinated tap water at 18 - 20°C and were fed twice weekly on blowfly larvae.

Triturus cristatus carnifex were obtained from Gerrard and Haig Ltd. Plethodon cinereus cinereus were a gift from Prof. H. C. MacGregor, Department of Zoology, University of Leicester.

BACTERIA AND PLASMID STOCKS

E. coli HB101 (Boyer et al, 1969) and E. coli HB101 containing the plasmid pBR322 (Bolivar et al, 1977) were obtained from Dr. A. Colman. Plasmid pAT153 (Twigg and Sherratt, 1979) was a gift from Dr. P. Slocombe and was transformed into E. coli HB101 as described below. E. coli χ 1776 was obtained from Dr. S. B. Primrose, and was transformed with pBR322 by the method of Norgard et al, (1978). E. coli MRC-1 was obtained from Dr. S. Brenner.

E. coli HB101 was maintained on N agar (see below). E. coli χ 1776 was maintained on supplemented L-agar (see below). E.coli MRC-1 was maintained on NY agar (see below). Long term stocks were kept at -20°C in a suspension of the relevant medium and sterile glycerol (1:1, v/v) on autoclaved glass beads. Bacterial stocks transformed with pBR322 or pAT153, or recombinants were maintained in the relevant medium containing 100 μg/ml ampicillin.

MEDIA

A.	N broth	Bacto nutrient broth	13 gm.
		Distilled water	to 1 litre
B.	L broth	Bactoyeast extract	5 gm.
		Bactotryptone	10 gm.
		NaCl	10 gm.
		Glucose	0.5 gm.
		Distilled water	to 1 litre

For plates add 15 gm. Difco Agar to 1 litre of medium.

Autoclave. Cool. Pour.

C.	M9 broth	Dissolve in order :	
	1) 10 x M9 salts	Na_2 HPO4	60 gm.
		KH_2 PO4	30 gm.
		NaCl	5 gm.
		(NH4)Cl	10 gm.
		Distilled water	to 1 litre

Autoclave.

2) Media	M9 salts	100 ml.
	Distilled water	760 ml.

Autoclave. Cool.

Autoclave the following separately, add in order, mix well before subsequent additions.

	20% w/v glucose	20 ml.
	0.1M MgSO4	10 ml.
	0.01M $CaCl_2$	10 ml.
	15% casaminoacids	100 ml.
(filter sterilise)	20 mg/ml thiamine	1 ml.

MEDIA

A.	N broth	Bacto nutrient broth	13 gm.
		Distilled water	to 1 litre
B.	L broth	Bactoyeast extract	5 gm.
		Bactotryptone	10 gm.
		NaCl	10 gm.
		Glucose	0.5 gm.
		Distilled water	to 1 litre

For plates add 15 gm. Difco Agar to 1 litre of medium.

Autoclave. Cool. Pour.

C.	M9 broth	Dissolve in order :	
	1) 10 x M9 salts	Na_2 HPO4	60 gm.
		KH_2 PO4	30 gm.
		NaCl	5 gm.
		(NH4)Cl	10 gm.
		Distilled water	to 1 litre

Autoclave.

2) Media	M9 salts	100 ml.
	Distilled water	760 ml.

Autoclave. Cool.

Autoclave the following separately, add in order, mix well before subsequent additions.

	20% w/v glucose	20 ml.
	0.1M MgSO4	10 ml.
	0.01M $CaCl_2$	10 ml.
	15% casaminoacids	100 ml.
(filter sterilise)	20 mg/ml thiamine	1 ml.

SUPPLEMENTED MEDIA

1) Supplemented L-broth for *E. coli* X1776

Basic L-broth		
Diaminopimelic acid	100 μg / ml	Final concentration
Thymidine	20 μg / ml	" "
d-Biotin	1 μg / ml	" "

2) NY broth for *E. coli* MRC-1

Basic L-broth		
Diaminopimelic acid	500 μg / ml	Final concentration
N-Acetylglucosamine	200 μg / ml	" "
Thymidine	40 μg / ml	" "

PREPARATION OF DNA

1. Axolotl DNA

Animals were anaesthetised in 0.1% n-ethyl aminobenzoate (MS222, Sigma) and blood was removed using a heparin coated pipette after cardiac puncture. The blood was suspended in five volumes of 100mM Nacl; 50 mM EDTA pH8.0 (SE Buffer) containing 5 µl of 1mg./ml. Heparin for each 100 ml. of SE buffer. The total volume was then measured and 20% sodium dodecyl sulphate (SDS) added to 0.5% final concentration. The solution was incubated with shaking for 30 minutes at 37°C.

An equal volume of phenol:chloroform:isoamyl alcohol:8 hydroxyquinoline (25:24:1:0.4; v/v/v/w) was added and the mixture was shaken vigorously for 15 minutes. The mixture was then centrifuged at 2500 rpm, 30 minutes at 4°C, in a Mistral 6L centrifuge. The aqueous phase was removed and the phenol layer was re-extracted with SE. The aqueous phases were pooled.

Sodium acetate was added to 0.2M. Two volumes of ice cold ethanol were added and the mixture was gently swirled on ice to precipitate the nucleic acids. The fibres were wound out of solution and washed three times in cold 70% ethanol. The precipitate was redissolved in 1 mM Tris pH7.5, 1 mM Nacl; 0.1mM EDTA (0.1 x TNE) overnight at 4°C with gentle shaking.

Ribonuclease A (RNAase A, Sigma) was added to 100 µg / ml and the mixture was incubated at 37°C for 2 hours. The solution was adjusted to 1 x TNE and 0.5% SDS and 100 µg / ml. pronase (Sigma). The mixture was again incubated for 2 hours at 37°C, then phenol extracted as above and the DNA precipitated. The precipitate was placed directly in a dialysis bag along with 1 - 5 mls of 10 mM Tris pH7.5; 1 mM EDTA (TE) and dialysed against TE with 3 changes of 1 litre over 36 hours at 4°C. The dialysis bag was preboiled in 50 mM

EDTA, washed thoroughly in distilled water and stored at 4°C in sterile distilled water.

The optical density of the DNA was measured. An absorbence of 1 at 260 nm. was taken as 50 μg/ml.

2. E. coli DNA

E. coli HB101 was grown to late log phase, chilled on ice and centrifuged at 6000 rpm in a 6 x 250 rotor in the MSE HS18 at 4°C. Cells were resuspended in 50 mM Tris pH8.0 and respun as above. The cells were resuspended in a minimum volume of 25% sucrose; 50 mM Tris pH8.0; 5 mM $MgCl_2$; 4 mg /ml lysozyme. The cells were incubated at 37°C for 30 minutes. EDTA (250 mM in 50 mM Tris pH8.0) was added to 30 mM final concentration. After 5 minutes on ice an equal volume of 0.5% Triton X-100; 50 mM Tris pH8.0 was added. The mixture was incubated for 30 minutes to 1 hour at 37°C with shaking to lyse the cells completely. The mixture was centrifuged at 17000 rpm, 4°C for 1 hour in the 8 x 50 rotor in an MSE HS18 centrifuge. The supernatant was decanted and an equal volume of phenol/chloroform/isoamyl alcohol added. The mixture was shaken and centrifuged for 15 minutes at 4°C in the MSE 6L at 4000 rpm. The supernatant was adjusted to 0.2M sodium acetate and two volumes of 95% ethanol added. Nucleic acid was precipitated overnight at -20°C and pelleted at 12000 rpm, 30 minutes at 4°C. The pellet was washed in 70% ethanol vacuum dried and redissolved in 1 ml. 1 x TNE. 100 μg RNAase A was added and incubated at 37°C for 1 hour, then phenol extracted, precipitated, washed and dried as above. The DNA was dissolved at 1 mg/ml in 1 x TNE.

For most purposes the DNA was used as above. However, in some cases the DNA was further purified by CsCl centrifugation. DNA was suspended in saturated CsCl to give an initial buoyant density of 1.710 gm/cc. at room temperature, and was centrifuged at 42000 rpm for 40 hours at 20°C in a

10 x 10 Titanium fixed angle rotor. Approximately 25 fractions were collected by puncturing the bottom of the centrifuge tube. Fractions containing the DNA were identified by their absorbence at 260nm . The relevant fractions were pooled and extensively dialysed in 0.1 x TNE at 4°C for further use.

3. Plasmid DNA

Large scale plasmid preparations were done as Colman et al (1978) with modifications. An overnight culture was used to seed the main culture using 1 ml for each 100 ml. of main culture. The main culture was grown to an A_{600} of 0.9 Chloramphenicol was added to a final concentration of 175 μg. / ml. to amplify the plasmid DNA and further incubated overnight at 37°C.

Cells were chilled on ice and pelleted by centrifugation at 6000 rpm, 30 minutes at 4°C in a 6 x 250 rotor in the MSE HS18. The cells were resuspended in 1/10 initial volume of 25 mM Tris pH8.0. The cells were recentrifuged as above and resuspended in 1/200 initial volume of 25% sucrose (w/v); 5 mM $MgCl_2$; 25 mM Tris pH8.0; 20 mg. / ml. lysozyme. The mixture was incubated at 0°C for 5 minutes. EDTA (250 mM in 50 mM Tris pH8.0) was added to a final concentration of 62.5 mM and incubated for a further 5 minutes at 0°C. Finally an equal volume of 0.5% Triton X-100; 50 mM EDTA; 25 mM Tris pH8.0 was added. The solution was incubated at 0°C for 20 minutes with occasional gentle rocking. The solution was then centrifuged at 17500 rpm for 1 hour, 4°C in an 8 x 50 rotor in the MSE HS18. The supernatant was poured off. This "cleared lysate" was made 0.2M sodium acetate pH7.5; 0.5% SDS and phenol extracted.

The aqueous phase was collected and 2.5 volumes of ice cold ethanol added. The nucleic acids were precipitated at -70°C, 30 minutes or -20°C overnight. The precipitated DNA and RNA was pelleted by centrifugation and resuspended in 1.1 ml water. Then 8.9 ml. of 9M Urea; 0.875% SDS;

0.27M sodium phosphate pH6.8 (Lysing buffer) was added.

A 10 ml. (approximate bed volume; height 1 cm., diameter 3 cm.) hydroxylapatite (DNA grade; Bio Rad) column was prepared by suspending the hydroxylapatite in 8M Urea; 0.24M sodium phosphate pH6.8 (Wash buffer). The column was packed under air pressure (4 lbs./square inch) and a sample of the run through volume retained as a pre-loading absorbence reference.

The DNA solution was applied to the column under air pressure and the column was washed with 80 - 100 mls. of wash buffer at approximately 5 mls./min. until the A_{260} of the eluate was below zero as compared to the preloading absorbence reference. The column was then washed with 80 mls. 0.01M Na Phosphate and DNA was eluted with 0.3M Na Phosphate. 1 ml. samples were taken and the A_{260} monitored against water as reference. Peak fractions were pooled, dialysed overnight against 0.1 x TNE precipitated with ethanol, and resuspended in 0.1 x TNE.

The chromatography step was usually performed in under an hour and several samples could be processed at the same time. The plasmid produced was predominantly supercoiled and was $\geqslant$ 95% pure of chromosomal DNA and completely free of RNA.

4. Small scale plasmid DNA preparations

For rapid screening of a large number of colonies the following procedure was obtained from I. Jones and was originally from Prof. D. Sherratt (University of Glasgow).

A 1 ml. culture of N broth and ampicillin (100 μg/ml in H_2o) in a sterile Bijou bottle was incubated overnight at 37°C in an orbital shaker. 0.5 ml. was transferred to 10 ml. N broth and Amp. and grown with shaking for 6 hours at 37°C. Chloramphenicol was added to 175 μg/ml and shaking was continued overnight at 37°C. The cells were centrifuged at 4000 rpm in a Beckman "Chillspin" at 4°C. All subsequent steps were performed on ice.

The medium was decanted, the pellet drained and resuspended by adding 200 µl 25% sucrose; 50 mM Tris pH8.0. The suspension was removed to a 1.5 ml Eppendorf microcentrifuge tube and 50 µl of lysozyme (20 mg/ml in 250 mM Tris pH8.0) added. The sample was left 5 minutes 0°C. Then 100 µl EDTA (250 mM in 50 mM Tris pH8.0) was added, the solution swirled gently with the micropipettor tip and left at 0°C for 10 minutes. 400 µl Triton X-100 (0.5% in 25 mM EDTA, 50 mM Tris pH8.0) was added and mixed gently as above. The solution was left 20 minutes, 0°C.

The solution was then spun in an Eppendorf microcentrifuge for 15 minutes. The pellet was removed through the supernatant as follows. A Gilson micropipettor (P-200) was set at 200 µl and the thumb depressor pressed almost all the way down. The tip was passed through the supernatant into the pellet. The thumb depressor was released and part of the pellet drawn into the tip. The pellet could then be drawn up through the supernatant and completely removed. 75 µl of 1M LiCl : 10% SDS was added to the supernatant and mixed. 800 µl of phenol/chloroform (1.1) was added, the solution was vortexed and centrifuged in a microcentrifuge. The phenol was removed through the aqueous phase and fresh phenol/chloroform added. The extraction was repeated and the phenol again removed. An equal volume of ether was added, vortexed and centrifuged. The lower aqueous phase was removed and nitrogen gas blown through to remove traces of ether. The supernatant was then split into two for convenience and 3 volumes 95% ethanol added. The DNA was precipitated at -70°C for 1 hour, centrifuged and the ethanol removed. The DNA was vacuum dried and resuspended in 0.1XTNE. An aliquot of each prep. was electrophoresed on agarose gels (0.8%). The rest was stored at -20°C DNA prepared this way could be digested by restriction endonucleases if required. The presence of RNA in the preparation obscures small fragments. This can be removed by incubation with ribonuclease A (50 µg/ml final concentration; 15 minutes 37°C).

Twelve colonies can conveniently be processed at the same time and two sets of 12 can be worked through in a day.

BACTERIAL TRANSFORMATION

A single colony of E. coli was taken up in 10 mls. of L broth and shaken overnight at 37°C. After overnight growth 1 ml. of culture was diluted into 50 ml. prewarmed L broth. The culture was grown to an A_{600} of 0.6 and centrifuged at 4000 rpm in an MSE "Chillspin" for 10 minutes at 4°C. The bacteria were resuspended in 20 mls. of 50 mM $CaCl_2$ and chilled on ice for 30 minutes. The cells were then recentrifuged as above and resuspended in 5 ml. of 50 mM $CaCl_2$ at a final density of approximately 2×10^9 cells/ml.

200 µl. of cells and up to 100 ng. of transforming DNA in 50 mM $CaCl_2$; 10 mM Tris pH7.5. The mixtures were left on ice for 60 minutes and were then transferred to 42°C for 3 minutes. 2.7 mls. of prewarmed L broth was added and incubated at 37°C for 60 minutes. 100 µl. aliquots were plated out onto selective medium. Ampicillin resistant bacteria were selected on agar containing 100 µg./ml. ampicillin; tetracycline resistant bacteria were selected on agar containing 12.5 µg./ml. tetracycline. Plates were incubated at 37°C.

METAPHASE CHROMOSOMES

Metaphase chromosomes were prepared from the dividing epithelial cells of the axolotl gut. Cell division in the epithelium was stimulated by intense feeding for two days, after which the animals were injected with 10 mg. of colchicine in 1 ml. of sterile distilled water and left for 48 hours.

The gut was removed, split open along its whole length and washed for 9 minutes in several changes of sterile distilled water. This step acted as a hypnotic treatment. After washing the gut was fixed with freshly made

ice-cold '3:1' (three parts Ethanol to 1 part Glacial Acetic Acid), after which it was divided into duodenum, mid gut and hind gut.

Squash preparations for *in situ* hybridisation were prepared as follows. Small pieces of tissue were removed from '3:1', placed in a drop of 45% Glacial Acetic Acid on a microscope slide and epithelial cells were scraped onto the slide and squashed between the slide and a siliconised cover-slip in the usual way. The coverslips were removed by the dry ice method and the slide washed in 95% ethanol and air dried. Prior to squashing the slides were covered in a thin film of gelatin in order to encourage retention of chromosomes to the slide. The slides were heated at 65°C overnight prior to hybridisation.

LAMPBRUSH CHROMOSOMES

Lampbrush chromosomes for *in situ* hybridisation were prepared as described by Callan and Lloyd (1960) and Hennen *et al* (1975) with the following modifications. The lampbrush observation chamber was constructed from 1 mm. thick plastikard attached to a glass microscope slide with double sided sellotape. The animals were anaesthetised in 0.1% MS222. Part of the ovary was removed and kept on ice in a dry, sealed embryo dish. The germinal vesicles were isolated in a solution consisting of 5 parts of 0.1M KCl and 1 part of 0.1M NaCl ('5:1'). Before removal of the nuclear envelope the nucleus was washed in dispersing medium and then transferred to the observation chamber to have its envelope removed. The saline used as the dispersing medium was 0.7 strength '5:1' containing 0.1% formaldehyde. The slides were left for up to 2 hours to allow the nuclear sap to disperse and were then centrifuged at 3000 rpm for 30 minutes at 4°C and fixed in cold 70% ethanol overnight, then washed in 95% ethanol, the plastikard removed and the preparation air dried.

The lampbrush preparations were made from oocytes of 1.5 - 1.7 mm.

diameter. Slides containing metaphase chromosomes, or lampbrush chromosomes, were incubated for 2 hours at 37°C in 1 x SSC containing 100 µg/ml RNAase A and 100µg/ml. RNAase T_1. After incubation the slides were washed in two changes of 2 x SSC, 5 minutes each at room temperature, two changes of 70% ethanol, 5 minutes each at room temperature, two changes of 95% ethanol, 5 minutes each at room temperature and air dried.

Chromosomal DNA was denatured either by heating to 80°C in 50% Formamide, 4 x SSC pH7.2 for 10 minutes or by immersing the slide in freshly made 0.07M NaOH for 3 minutes at room temperature. After this the slides were washed and dehydrated as above.

DNA - RNA hybridisations were performed only on lampbrush chromosomes without pretreatment.

The ^{3}H-labelled probe DNA was denatured by boiling in distilled water for 5 minutes and then made 50% Formamide, 4 x SSC pH7.2 and used at approximately 2 x 10^5 cpm per slide in a volume of 20 µl. (Specific activity of probe DNA's ranged from 3 - 5 x 10^6 cpm/ µg).

The probe DNA was applied to the preparation and a coverslip was carefully lowered onto the liquid ensuring that no air bubbles were trapped between the slide and the coverslip. The coverslips were preboiled in 0.2M HCl, then extensively washed in running tap water, rinsed in distilled water then stored in 70% ethanol.

Hot candle wax was applied to the edges of the coverslips to prevent evaporation and the slides were incubated at 37°C for 18 - 24 hours. After hybridisation the slides were washed as follows : 2 x SSC at room temperature three changes of 5 minutes each; two changes of 2 x SSC preheated to 65°C, 5 minutes each; two changes of 2 x SSC at room temperature 5 minutes each; 5% TCA precooled to 4°C, 5 minutes; two changes of 2 x SSC at room temperature, 5 minutes each; three changes of 70% ethanol, 5 minutes each; two changes of 95% ethanol, 5 minutes each, air dried.

diameter. Slides containing metaphase chromosomes, or lampbrush chromosomes, were incubated for 2 hours at 37°C in 1 x SSC containing 100 µg/ml RNAase A and 100 µg/ml. RNAase T_1. After incubation the slides were washed in two changes of 2 x SSC, 5 minutes each at room temperature, two changes of 70% ethanol, 5 minutes each at room temperature, two changes of 95% ethanol, 5 minutes each at room temperature and air dried.

Chromosomal DNA was denatured either by heating to 80°C in 50% Formamide, 4 x SSC pH7.2 for 10 minutes or by immersing the slide in freshly made 0.07M NaOH for 3 minutes at room temperature. After this the slides were washed and dehydrated as above.

DNA - RNA hybridisations were performed only on lampbrush chromosomes without pretreatment.

The ^{3}H-labelled probe DNA was denatured by boiling in distilled water for 5 minutes and then made 50% Formamide, 4 x SSC pH7.2 and used at approximately 2 x 10^5 cpm per slide in a volume of 20 µl. (Specific activity of probe DNA's ranged from 3 - 5 x 10^6 cpm/ µg).

The probe DNA was applied to the preparation and a coverslip was carefully lowered onto the liquid ensuring that no air bubbles were trapped between the slide and the coverslip. The coverslips were preboiled in 0.2M HCl, then extensively washed in running tap water, rinsed in distilled water then stored in 70% ethanol.

Hot candle wax was applied to the edges of the coverslips to prevent evaporation and the slides were incubated at 37°C for 18 - 24 hours. After hybridisation the slides were washed as follows : 2 x SSC at room temperature three changes of 5 minutes each; two changes of 2 x SSC preheated to 65°C, 5 minutes each; two changes of 2 x SSC at room temperature 5 minutes each; 5% TCA precooled to 4°C, 5 minutes; two changes of 2 x SSC at room temperature, 5 minutes each; three changes of 70% ethanol, 5 minutes each; two changes of 95% ethanol, 5 minutes each, air dried.

The slides were coated with Kodak NTB-2 nuclear track emulsion diluted 1+1 with distilled water, and stored in light-tight boxes containing silica gel as desiccant.

The autoradiographs were developed in freshly made D-19 developer for 2½ minutes at room temperature, rinsed in distilled water, fixed in kodafix for 5 minutes, thoroughly washed in distilled water and stained with Giemsa. The coating with emulsion and developing were performed under safelight conditions using a 15W bulb and a Kodak 2B red filter. Exposure time varied from three to six weeks.

All cytological preparations used for *in situ* hybridisations were prepared on slides obtained from the American manufacturers Thomas and Co. (Hennen *et al*, 1975).

DETERMINATION OF C VALUES BY FEULGEN MICRODENSITOMETRY

The C value of *A. mexicanum* was determined in erythrocyte nuclei with reference to either *Triturus cristatus carnifex* or *Plethodon cinereus cinereus* erythrocyte nuclei stained on the same slide. Feulgen dye content of individual nuclei was determined by photometric measurements using the Vickers V85 Integrating Microdensitometer at a wavelength of 560 nM.

Animals were anaesthetised in 0.1% MS222 and blood was removed from a cardiac puncture using a heparinised pipette. In the case of *P. cinereus cinereus* the anaesthetised animals were decapitated and blood was applied directly from the wound to the slide. Blood from one animal was smeared on one half of a slide and from a second animal on the second half, so that differences between slides should not affect comparisons between animals. The slides were allowed to air dry, were fixed in '3:1', washed in two changes of 95% ethanol, 5 minutes each and air dried.

Preparations were stained with Feulgen dye as follows. Slides were rinsed with distilled water, drained and placed in freshly made 5M HCl at 20°C for 20 minutes. The timing of hydrolysis was found to be critical for

good staining. The slides were then washed in distilled water, three changes of 5 minutes each. Slides were placed in leuco basic fuschin for 90 minutes at room temperature in a Coplin jar covered in tin foil. The fuschin was prepared as follows :

A 0.5% solution (w/v) of basic fuschin was made by dissolving the solid in hot distilled water. The solution was allowed to cool and was then filtered. One tenth volume of fresh 1M HCl was added. Then sodium metabisulphate was added at 1 g. per 100 mls. of basic fuschin. The solution was allowed to stand in the dark overnight. The solution was decolourised by adding 1 g. charcoal per 200 mls. and leaving for 2 hours in the dark in a fumehood. The solution was then filtered. This process was repeated until the solution was decolourised. The solution was stored at $4^{\circ}C$ in a dark bottle.

After staining, the slides were washed 3 x 5 minutes at room temperature in "SO_2 water". "SO_2 water" was prepared immediately before use by mixing 10% sodium metabisulphite, 5M HCl and distilled water (1:1:20). Slides were then washed in distilled water, 5 minutes, 50% ethanol, 5 minutes 70% ethanol, two changes of 5 minutes each, 95% ethanol, two changes of 5 minutes each, and air dried.

ENZYMES

A. Restriction Endonucleases

Restriction endonucleases were obtained from Uniscience, Cambridge; Miles or Boehringer Mannheim, or were prepared in the laboratory by a method provided by Prof. D. Sherratt, University of Glasgow.

Digestion buffers were those recommended by the manufacturer. Where digestion with two different enzymes was required the enzyme requiring more stringent conditions was added after total digestion with the first enzyme, and digestion conditions were adjusted to meet the requirements

of the second enzyme. Where two enzymes required the same digestion conditions they were usually added together.

Usually a twofold excess of enzyme was used and digestion was allowed to proceed for $\geqslant$ 2 hours, at the recommended temperature.

B. SI Nuclease

SI Nuclease from A. oryzae was purchased from Sigma.

Reactions were performed in 230 mM NaCl; 0.5 mM $ZnCl_2$; 30 mM sodium acetate pH4.5 (Schacht and Hogness, 1973). The NaCl was added to prevent digestion at the ends of duplexes caused by DNA "breathing" in low salt. Digestions were performed for 1 hour at 37°C with sufficient enzyme to degrade $\geqslant$ 95% of single stranded DNA while not degrading double stranded DNA.

Reactions were assayed by binding aliquots to DE81 paper in 0.15M sodium phosphate buffer pH7.6, and washing digestion products with 5 x 5 minute washes of 0.15M sodium phosphate buffer.

C. DNA polymerases

Kornberg DNA polymerase was purchased from Boehringer Mannheim.

Reactions were performed in 5mM $MgCl_2$; 10mM 2-mercaptoethanol (β-ME); 50 mM Tris pH7.5; 100 μg/ml Bovine Serum Albumin (BSA). The concentrations of nucleotide triphosphates were varied as required. Reactions were performed at 15°C to prevent the 'strand-hopping' activity of DNApolymerase (Rigby et al, 1977). Reactions were terminated, usually after 90 minutes, by phenol-chloroform extraction and Sephadex G -50 chromatography to remove unincorporated nucleotide triphosphates.

Klenow Polymerase - the large subunit of E. coli DNA polymerase, was obtained from Dr. N. Mann or was a gift from Dr. P. Turner, and was used solely for 3' end labelling.

D. T4 DNA Ligase

Phage T4 DNA Ligase was purchased from Bethesda Research Laboratories and was from T4 am N82 infected E. coli cells.

Reactions were performed in 6.6 mM $MgCl_2$; 1 mM ATP; 66 mM Tris pH7.6. Ligations were carried out at 4°C for 48 hours. DNA concentrations and amounts of Ligase used were varied to produce efficient ligation conditions (Dugaizyk et al, 1975).

RADIOACTIVE LABELLING OF DNA

(1) in vivo labelling

Axolotl embryos (early neurulae) were manually dejellied in 1 x Barth-X and micro-injected with 80 nCi each of ^3H thymidine into the abdominal regions by Dr. H. R. Woodland. After 24 hours at 15°C the embryos were homogenised in a 1 ml. homogeniser (heat treated and autoclaved) in 1 x SE buffer. ¼ volume of 5M $NaClO_4$ and 1/20 volume of 10% SDS were added and the suspension was incubated for 30 seconds at 37°C to lyse the cells. The procedure for extraction of blood cell DNA was then followed exactly.

It was noted that the majority of the thymidine leaked out of the embryos during the incubation. Approximately 80% of the injected label could be located in the Barth-X after 24 hours at 15°C. Specific activity was 1.5×10^3 cpm/μg.

(2) in vitro labelling

i) Nick translation

DNA was labelled by "Nick translation" by a modification of the method of Rigby et al, (1977). For ^3H labelling an equimolar amount of each ^3H deoxynucleotide triphosphate (dNTP) was mixed and 20 - 40 μCi of this mix was used per reaction. The ^3H dNTPs were obtained in 50%

aqueous ethanol (Amersham) and were vacuum desiccated before use.

For ^{32}P labelling the ^{32}P dNTPs were either dGTP or dCTP. These were obtained as a stable aqueous solution at 10 mCi/ml., 2000 Ci/m mole (Amersham). 25 μCi of one ^{32}P dNTP and a twofold molar excess (25 pmoles) of the three cold dNTPs were used in a standard reaction. The standard reaction conditions contained, in addition to the dNTPs, 0.05-1 μg of DNA, 10 pg of DNAase I for each 100 ng DNA added and 5 - 10 units of Komberg DNA polymerase I in 50 mM Tris pH7.8 5mM $MgCl_2$, 10 mM mercaptoethanol in a final volume of 20 μl.

The reaction mix was set up without the DNA polymerase and incubated at 37°C for 15 minutes to activate the DNAase I. The DNA polymerase I was then added and the mix transferred to 15°C for 3 hours. At 15°C the "strand-hopping" action of DNApolymerase I is reduced (Rigby et al, 1977). The reaction was terminated by diluting the solution to 100 μl and adding 1/10 volume 1M LiCl/10% SDS. 10 μg of E. coli tRNA was added and then an equal volume of phenol/chloroform/iso-amyl alcohol. The mixture was vortexed, centrifuged for 5 seconds in a microcentrifuge and the supernatant was removed. The phenol layer was re-extracted with 100 μl of 0.1 x TNE. The two supernatants were pooled and passed down a 2 ml. Sephadex column. The exclusion peak was collected and precipitated in ethanol if required.

Incorporation was usually 25 - 50% of the input radio label, and specific activities ranged from 5×10^6 - 3×10^8 dpm/μg.

ii) End labelling

DNA was end labelled by the incorporation of ^{32}P dNTP into an exposed staggered 5' end after restriction endonuclease treatment.

1 - 5 μg DNA was restricted with the specific restriction enzyme, phenol extracted, precipitated with ethanol and redissolved in Klenow reaction buffer (see earlier) 10 - 20 μCi of the required dNTP and 1 unit

of Klenow fragment of Polymerase I were added in a final volume of 10 µl. The reaction was allowed to proceed for 3 hours and then was processed exactly as for nick translation. Specific activities of 2 x 10^5 - 10^6 Cerenkov cpm/µg were achieved.

SHORT DRIVER - LONG TRACER REACTIONS : PREPARATION OF SIZE FRACTIONATED ^3H LABELLED DNA.

1. Preparation of tracer DNA

^3H genomic DNAs of various single strand sizes were prepared as follows. Incubations were set up according to the nick translation method described above. The amount of DNAase I and the incubation time were varied to produce fragments of variable single strand size after denaturation. The nick translation products were pooled and passed over Sephadex G-50, made 0.2M Na Acetate and ethanol precipitated. The DNA was washed in 70% ethanol, vacuum dried and redissolved in Bam HI restriction buffer. This was done to allow comparison of migration with the restriction endonuclease treated plasmids used as size markers. It has been noted that migration during electrophoresis is dependent upon the salt present in the sample.

2. Size fractionation of labelled DNA

Size fractionation was done by electrophoresing denatured DNA through neutral agarose gels exactly as in McDonnell et al, (1977). To test the efficacy of the procedure several trials were done. Fig. 2.1 shows the results of one such trial separation.

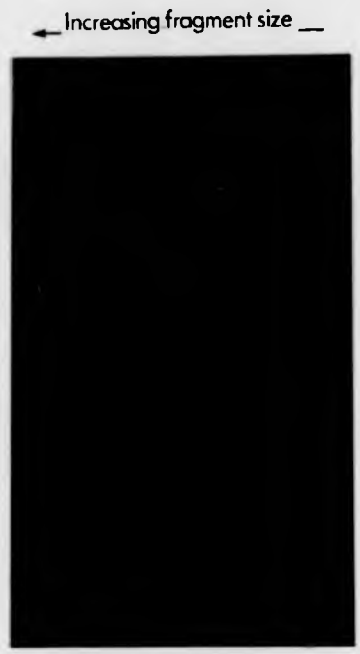
For the purposes of this test the DNA was cold, sheared by sonication and was double stranded to allow easier visualisation using ethidium bromide. A neutral agarose gel was cast with two sets of slots, one set parallel to the first dimension of electrophoresis. After size fraction-

LEGEND TO FIG. 2.1

SIZE FRACTIONATION OF SHEARED AXOLOTL DNA

DNA was sheared and separated into size classes as described in the text. After elution from hydroxylapatite the DNA was diluted with H_2O to 0.1M Na phosphate and electrophoresed on a neutral agarose gel in a Tris phosphate buffer system (see text).

The size distribution of the separated fractions ranges from $\geqslant$ 4 kbp to around 500 bp, for this particular trial.


Increasing fragment size

ation the second slots were filled with hydroxylapatite and the gel was rotated 90° and the DNA was electrophoresed into the HAP. This was then removed and the DNA was eluted at 60°C with 0.4 MNaPB, pH6.8.

Other tests using ^{32}P labelled single stranded DNA to assay the progress of single strands through the procedure produced effectively the same pattern of fractionation (data not shown).

For the major experiments the DNA was nick translated using all four tritiated (^{3}H) dNTPs to reduce spurious effects which might be caused by the various G+C contents of genomic subclasses, and to provide probes with longer life times.

For the main fractionation the gel was run without ethidium bromide present and the markers were cut off the gel and stained separately. The relative migration was noted and graphed vs. the second slot positions. The size of each tracer given in the results is the midpoint of the relevant slot.

AGAROSE GEL ELECTROPHORESIS

(1) Horizontal gels

Agarose gels of varying concentration (0.75 - 3%) were prepared by heating, varying amounts of agarose (Sigma, Type II) in 100 mls. of TEA buffer (TEA = 40 mM Tris, 1mM EDTA, 5 mM Na acetate pH7.8). The gels were degassed, cooled to 50°C and ethidium bromide added to 1 µg/ml. before pouring. Gels were cast on a horizontal 15 x 20 cm. glass plate. Sellotape was placed around the glass plate to seal the plate and prevent leakage. Slot formers of various dimensions were used to provide a variable number of slots per gel as required.

Gels were electrophoresed in TEA buffer + 1 µg/ml. ethidium bromide either at 30v for 16 hours or 100v 3 hours. They were then removed and photographed using either Polaroid Type 665 Land Film with a yellow filter,

← Increasing fragment size →


Increasing fragment size

ation the second slots were filled with hydroxylapatite and the gel was rotated 90° and the DNA was electrophoresed into the HAP. This was then removed and the DNA was eluted at 60°C with 0.4 MNaPB, pH6.8.

Other tests using ^{32}P labelled single stranded DNA to assay the progress of single strands through the procedure produced effectively the same pattern of fractionation (data not shown).

For the major experiments the DNA was nick translated using all four tritiated (^{3}H) dNTPs to reduce spurious effects which might be caused by the various G+C contents of genomic subclasses, and to provide probes with longer life times.

For the main fractionation the gel was run without ethidium bromide present and the markers were cut off the gel and stained separately. The relative migration was noted and graphed vs. the second slot positions. The size of each tracer given in the results is the midpoint of the relevant slot.

AGAROSE GEL ELECTROPHORESIS

(1) Horizontal gels

Agarose gels of varying concentration (0.75 - 3%) were prepared by heating, varying amounts of agarose (Sigma, Type II) in 100 mls. of TEA buffer (TEA = 40 mM Tris, 1mM EDTA, 5 mM Na acetate pH7.8). The gels were degassed, cooled to 50°C and ethidium bromide added to 1 µg/ml. before pouring. Gels were cast on a horizontal 15 x 20 cm. glass plate. Sellotape was placed around the glass plate to seal the plate and prevent leakage. Slot formers of various dimensions were used to provide a variable number of slots per gel as required.

Gels were electrophoresed in TEA buffer + 1 µg/ml. ethidium bromide either at 30v for 16 hours or 100v 3 hours. They were then removed and photographed using either Polaroid Type 665 Land Film with a yellow filter,

or Kodak Pan X with a red filter, over a U.V. transilluminator.

In special circumstances, the gel running buffer was phosphate based and was 89 mM Tris pH8.3; 2.5 mM EDTA; 23 mM Na Phosphate.

(2) Vertical tube gels

Polycarbonate cylinders internal diameter 8 mm. and length 100 mm. were sealed at one end with Saran Wrap held in place with an elastic band. Agarose and Ethidium bromide was prepared as above and was poured into each tube. When set the gel was pushed out 1 cm. from below and the gel sliced level with the top of the tube. The gel was then allowed to slide back into the tube, so forming a loading well at the top of the gel. The Saran Wrap was pierced with holes to allow contact with the running buffer. The gels were attached to the vertical assembly and running buffer added to both top and bottom chambers. The gels were run at 40 mA (constant current) for 2 hours. Gels were photographed as above.

HYBRIDISATION PROCEDURES

i) Solution hybridisations : Cot analysis

High molecular weight DNA was sheared to 500 bases (single stranded weight average) by sonication in a Dawes Soniprobe (1mA; 10 x 30 secs. with 30 secs. gap between bursts, on ice), DNA was sheared to 2.5 kb (single stranded weight average) by 2 x 30 secs. bursts with a 30 sec. gap, on ice. The DNA was passed over a small Chelex-100 column, pre-equilibrated with TE, in a Pasteur pipette. DNA was then made 0.2 M Na Acetate pH7.5 and precipitated with 2 volumes of cold 95% ethanol. The DNA was redissolved in a small volume of TE.

Before reassociation DNA was made 0.12M Na phosphate pH6.8 (equal volumes of $Na_2 HPo_4$ and $Na H_2Po_4$, molarity refers to the phosphate group). The solution was boiled 5' at 100°C. 4.8M Na phosphate (as above) or

distilled water were added where required to adjust the phosphate concentration for reassociation.

The solutions were reassociated at a criterion equivalent to 60°C, 0.12M Na phosphate (see Britten et al, (1973) for the effect of $[Na^+]$ on rate and Appendix II, this thesis, for the effect of $[Na^+]$ on Tm from which corrections to Equivalent Cot may be made).

At various times 50 µg aliquots were removed, made 0.12M Na Phosphate and 1 ml. in volume and cooled on dry ice/acetone to quench the reaction. Samples were stored at -20°C before fractionation.

New lots of DNA grade Hydroxylapatite (HAP) were assayed as follows : Specificity of elution of double and single stranded DNA's were measured by loading trace amounts of P^{32} labelled "nick translated" DNA, either double or single stranded, together with 50 µg sheared axolotl DNA onto a 1 ml. bed of HAP at 60°C. A jacketed column maintained at 60°C was used. Column temperature was measured by placing a thermometer in the solution above the HAP before elution. Elution was carried out under positive air pressure. Standard elution procedure was 3 x 3 ml. washes with 0.12M Na phosphate, to elute single stranded material and 3 x 3 ml. 0.4M Na phosphate to elute double stranded material. A typical batch test is shown in Fig. 2.2.

The standard fractionation procedure was modified for the fractionation of the 2.5 kb reassociation products. Kiper (1978) has shown that long DNA may form complexes due to the presence of more than one repeated sequence per fragment length. These complexes are not efficiently eluted at 60°C by 0.4M Na phosphate, but may be eluted at 95°C in 0.12M Na phosphate. For the fractionation of reasssociation products of 2.5 kb DNA, the elution of duplex containing material was modified accordingly.

When cold DNA was fractionated the A_{260} of each wash was monitored with reference to the equivalent wash of a fractionation performed with

LEGEND TO FIG. 2.2

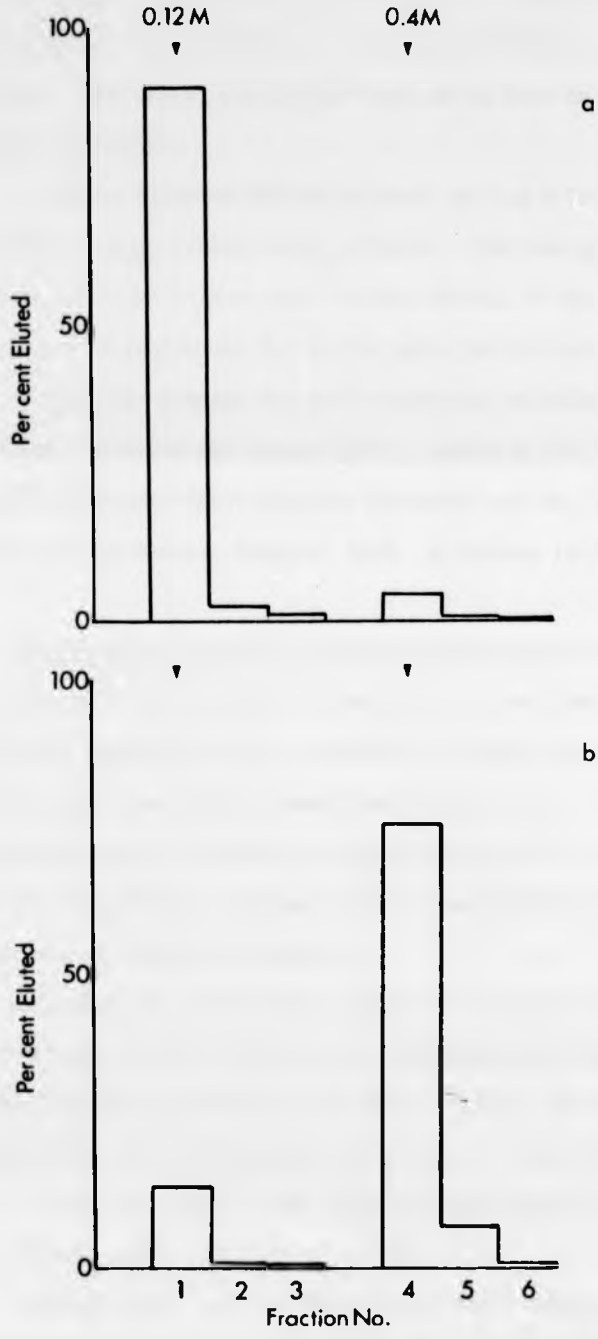
ELUTION OF DNA FROM HYDROXYLAPATITE AT 60°C

E. coli DNA was 'nick translated' with ^{32}PdGTP (Rigby *et al*, 1977) and was mixed with sheared, cold *E. coli* DNA. 50 µg of DNA in 1 ml. of 0.12M. Na phosphate buffer pH6.8 was loaded onto a 1 ml. hydroxylapatite column maintained at 60°C.

The DNA was either denatured prior to loading by boiling for 5 minutes at 100°C (a) or native (b).

The DNA was allowed to equilibrate at 60°C then was eluted by three 3 ml. washes with 0.12M Na phosphate pH6.8 (1 - 3) followed by three 3 ml. washes with 0.4M Na phosphate pH6.8 (4 - 6).

The fractions were counted by Cerenkov counting and are plotted as % total counts released *vs* fraction number.


0.12 M
0.4M
100
50
0
Per cent Eluted
a
100
50
0
Per cent Eluted
b
1
2
3
4
5
6
Fraction No.

no added DNA. DNA content in each wash was calculated on the basis of a hyperchromicity of 40% for single stranded DNA, i.e. an A_{260} of 28 = 1 mg/ml. The duplex portion was assumed to have an A_{260} of 20 for a 1 mg/ml. solution.

% single stranded DNA was plotted vs Log E Cot. Cot was conveniently measured as A_{260} (denatured)/$_2$ x hours. (Britten et al, 1973) reassociation in various salt concentrations at the same criterion were converted to equivalent Cot by the table in Britten et al, 1973.

A best fit through the data points was calculated using the computer programme of Kells and Straus (1977), modified for the University of Warwick Burroughs B6700 computer by myself and Mr. N. Coveney of the University of Warwick Computer Unit. A listing is given in Appendix II.

ii) Filter hybridisations : Saturation hybridisations

DNA in a final volume of 400 µl 0.1 x SSC (SSC = 0.15M NaCl, 0.015M Na acetate pH7.2) was denatured by adding an equal volume of 1M NaOH and incubating at room temperature for 30'. The DNA was neutralised by the addition of approximately 2 ml. 0.2M HCl; 9 x SSC; phenol red indicator. Enough solution was added to turn the indicator from pink to yellow/colourless.

The solution was passed through a nitrocellulose filter (HAWP, 13 mm, 0.22 µm pore size) which had been pretreated by wetting with distilled water followed by washing with 6 mls. 6 x SSC. The solutions were passed through the filter under negative pressure. After loading of the DNA the filters were washed with 6 mls. 6 x SSC, dipped in 70% ethanol, blotted and baked for 2 hours at 60°C.

Hybridisation was performed in 3 x SSC + Denhardts solution (0.02% of each of Ficoll, Polyvinyl pyrrolidone, Bovine Serum Albumin) at 65°C. Briefly the probe DNA was denatured made 3 x SSC + Denhardts and added to a filter, pre-wetted in 3 x SSC + Denhardts, in a

scintillation vial. A duplicate filter was added and paraffin oil was layered over the solution. The tubes were tightly stoppered and placed at 65°C for 24 - 36 hours. After hybridisation, the filters were washed in a large volume of 2 x SSC at room temperature. Three washes were performed for 20' each. The filters were dried at 60°C and counted in Toluene PPO/POPOP in a Packard scintillation counter.

iii) Southern Transfer and Filter Hybridisation

DNA was fractionated by size on agarose gels and then transferred to nitrocellulose paper by the method of Southern (1975). Briefly, the gel was soaked in 0.5M NaOH, 1.5M NaCl for 15 minutes, then transferred to 3M NaCl, 1M Tris pH5.0 for a further 15 minutes after which the DNA was transferred to nitrocellulose at room temperature for 18 hours. The filter was washed for 15 minutes in 2 x SSC, baked at 80°C for 2 hours and stored at room temperature until used.

Before hybridisation the filters were placed in 3 x SSC at 65°C for 30 minutes then 3 x SSC, 1 x Denhardts (Denhardt, 1965) for 3 - 16 hours. Radioactively labelled probe DNA in 0.1 x TNE was denatured by boiling for 10 minutes, then made 3 x SSC, 1 x Denhardt, 0.1% SDS. Hybridisation was performed in a sealed plastic bag at 65°C for 18 - 24 hours.

After hybridisation the filter was washed in three changes of 3 x SSC, 1 x Denhardt at 65°C for 40 minutes each, followed by a final wash in 3 x SSC at 65°C. The filters were then blotted dry and baked at 50°C for 2 hours.

The filters were marked with radioactive ink and exposed to Kodak X-Omat film either at room temperature without an intensifying screen or at -70°C with a screen. Usual practice was to develop the film after overnight exposure to gauge the optimum exposure time. Films were developed at room temperature in Kodak DX-80 (diluted 1+4 with water) and fixed in Kodak FX-40 (diluted 1+4 with water).

iv) Grunstein-Hogness hybridisations

This method is based on that of Grunstein and Hogness (1975). Millipore filters (0.45 μm pore) were cut to fit 9 cm. Petridishes and a grid pattern was drawn on in pencil. The filters were sterilised by exposure to short wave ultraviolet illumination, 5 minutes for each side.

The filters were placed on Lagar + ampicillin (for recombinants with insertion at the Bam HI site), and oriented. Colonies were pricked onto the surface of the nitrocellulose and onto a master plate. The colonies were grown on the surface of the nitrocellulose overnight at 37°C, (E. coli HB101 containing pBR322 were streaked as controls). The filters were then transferred to L agar + antibiotic plates containing 200 μg/ml chloramphenicol. This acts to amplify the plasmid present in each cell (see Plasmid preparation). The plates were returned to 37°C for 24 hours.

The bacteria were lysed in situ by the method of Humphries et al, (1978). A sheet of Whatman 3MM filter paper was placed on a glass plate and soaked in a 0.5M NaOH. The filters containing the colonies were laid on the paper, colony side up, for 10'. This lyses the bacteria and denatures the DNA.

The filters were then transferred to another sheet of 3MM paper soaked in 1M Tris pH7.4, for 5'. This was repeated until the pH of the solution in the Whatman remained at 7.4 when the nitrocellulose was added. The filter was then transferred to a sheet containing 1.5M NaCl 0.5M Tris pH7.4.

Colony debris was removed by transferring the filters to a clean 3MM sheet, overlaying with another 3MM sheet and "blotting" the filters with firm even pressure. When the top 3MM was removed the majority of the colony was removed leaving behind the imprint of the colony on the filter. The filters were then baked at 80°C for 2 hours, then stored at 4°C. Hybridisation of ^{32}P labelled probes was exactly as for Southern hybridisations.

DNA DENATURATION CURVES

1. Optical

High molecular weight DNA dissolved at 50 µg/ml in 0.12M Na phosphate pH6.8 was injected into the optical chamber of a Gilson Thermocuvette. 0.12M Na phosphate was injected into a parallel chamber as a reference blank. The DNA was heated at 0.5°C/min. using a Thermoprogrammer attached to the Gilson spectrophotometer, and change in absorbance was monitored with an automatic recorder. Hyperchromicity was measured as absorbance at 260 nm between 98°C and 60°C.

To obtain the Tm (point of 50% denaturation) the plot was rescaled setting the A_{260} 60°C at 0 and the A_{260} 98°C at 100% hyperchromicity respectively.

2. Hydroxylapatite

Radioactively labelled DNA was made 0.12M Na phosphate buffer pH6.8 (NaPB) and loaded onto a 1 ml. hydroxylapatite (HAP) column at 60°C. Where small amounts of DNA were to be melted the DNA was premixed with 50 µg sonicated E. coli DNA to reduce nonspecific binding to the HAP. The column was washed 3 x with 5 mls. 0.12M NaPB and the eluate collected. 5 mls. 0.12M NaPB was loaded above the column, a thermometer was inserted into the solution and the temperature was raised by the desired increment, usually 5°C. After 5 minutes, at the new temperature the column was washed with 2 x 5 mls. 0.12M NaPB, the first wash being the solution above the HAP during the temperature rise. The fractions were again collected. The process was repeated until the temperature of the column was 98°C. At this point the 2 x 5 ml. washes were followed by 2 x 5 ml. 0.4M NaPB and then the column was dissolved in 25% TCA (Kiper & Herzfeld, 1978). For ^{32}P the fractions were Cerenkov counted at 70% gain. For ^{3}H or $^{32}P,^{3}H$ mixes aliquots of each fraction were mixed with Triton-Toluene-

PPO-POPOP and counted at 70% gain for ^{3}H: 1.2% gain for ^{32}P.

The counts eluted at each temperature were summed and given as a percentage of the elutable counts bound at 60°C. Material not released by HAP at 98°C in 0.4M NaPB was usually less than 0.5% of the total and was not corrected for.

The elution profile was expressed either as cumulative percentage counts eluted vs temperature or counts eluted at each temperature vs temperature.

Tim (point of irreversible melting of 50% of the loaded material) was calculated from the cumulative curve.

Where melting curves were to be directly compared sheared ^{3}H labelled E. coli DNA was added as an internal standard.

CHAPTER 3

RESULTS AND DISCUSSION I

PHYSICAL PARAMETERS OF THE AXOLOTL GENOME

The purpose of this section is to provide data on various physical parameters of the axolotl genome which will serve as a framework for the subsequent studies on the repetitive fraction of the genome.

1. Genome size

Several estimates are available for the C value of *Ambystoma mexicanum*. These estimates have been reviewed by Bachmann (1971). The average of these determinations gives a value of 37 pg.

In view of the wide spread of published values it was decided to obtain an estimate of the genome size by microphotometric measurements of Feulgen dye contents of erythrocyle nuclei. To minimise errors in measurement a standard with a high C value was chosen. This allows comparison of nuclei of similar dimensions, so that variations in staining during the Feulgen process due to size and volume may be minimised. Both axolotl and standard blood samples were smeared on the same slide and therefore the subsequent processing was identical for each smear. Several preparations were made and a large number of individual nuclei (around 50 per smear per slide) were measured. It was then possible to determine a mean value for each slide, and for a group of slides.

Initially the standard used was *Triturus cristatus carnifex* (C = 23 pg, Varley *et al*, 1980 b; Horner pers. comm.). The results using this standard are given as Table 3.1A. The value of 38 pg agrees well with the consensus of the published values, and so was adopted as the C value for the main body of the thesis. Subsequently a small number of *Plethodon cinereus cinereus* (C = 20 pg, Mizuno & MacGregor, 1974) were obtained from Prof. H.C. MacGregor University of Leicester. A second series of estimations were made using blood from these animals as standard. The results are shown in Table 3.1B.

TABLE 3.1

C VALUE DETERMINATIONS

(a) CELL STANDARD TRITURUS C·CARNIFEX ERYTHROCYTES : C = 23pg [a]

SLIDE NO.	MEAN RELATIVE VALUE (AXOLOTL/STANDARD) [b]
1	1.67
2	1.65
3	1.61
MEAN	1.64
ABSOLUTE VALUE	37.7pg

(b) CELL STANDARD PLETHODON C·CINEREUS ERYTHROCYTES : C = 20pg [c]

SLIDE NO. [b]	MEAN RELATIVE VALUE (AXOLOTL/STANDARD) [d]
1	1.54
2	1.62
3	1.84
MEAN	1.67
ABSOLUTE VALUE	33.4pg

a. Varley et al, 1980b; Horner, pers. comm.

b. Mean relative value was the ratio of the mean value from at least 50 nuclei from both axolotl and standard smears.

c. Mizuno and MacGregor, 1974.

d. Mean relative value was the ratio of the mean value from at least 20 nuclei from both axolotl and standard smears.

It is immediately noticeable that there is a wide variation between different slides and the final value obtained is only 33 pg. One possible explanation for these results is that the Plethodon blood was applied to the slides as a droplet directly from the decapitated animal. In several cases clotting was noticeable and may have affected subsequent staining. Blood from Triturus and the axolotl was obtained from cardiac puncture, the blood being taken up in a heparinised pipette. For the subsequent examination of the genome the C value has been accepted as 38 pg based on 3.1A. The variation in the second series of measurements serves as a reminder of the need to keep all conditions identical when preparing the slides for measurement.

2. % G+C content

The % G+C content of the axolotl was measured by two methods. First, analytical ultracentrifugation in neutral CsCl gradients was performed and the buoyant density was measured relative to a standard DNA of known buoyant density, Micrococcus lysodeikticus ($\rho = 1.731$). The calculated buoyant density for the axolotl was 1.704 g/cc which corresponds to a % G+C content of 45% according to the empirical formula

$$\% \text{ G+C} = \left[\frac{\rho - 1.660}{0.098}\right] \times 100 \qquad \text{(Mandel and Marmur, 1971)}$$

Buoyant density centrifugation using E.coli DNA (50% G+C, $\rho = 1.710$) as internal standard indicated that the axolotl has a buoyant density which is similar to but slightly lighter than that of E. coli, which is consistent with the above data. Fig. 3.1 shows the microdensitometer scan of a photographic negative taken from an analytical ultracentrifuge run as described. From the densitometric scan shown in Fig. 3.1 it can be seen that the density peak for the axolotl is essentially symmetrical, no major satellite peaks are visible in neutral CsCl.

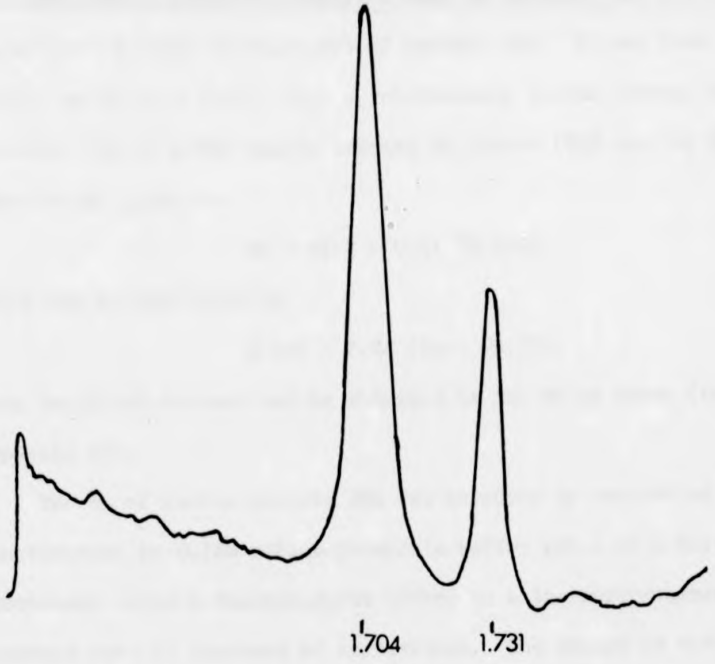
Thiery et al (1976) have performed ultracentrifugation analysis on the genomes of 25 eukaryotes, from mammals to yeast. They have noted phylo-

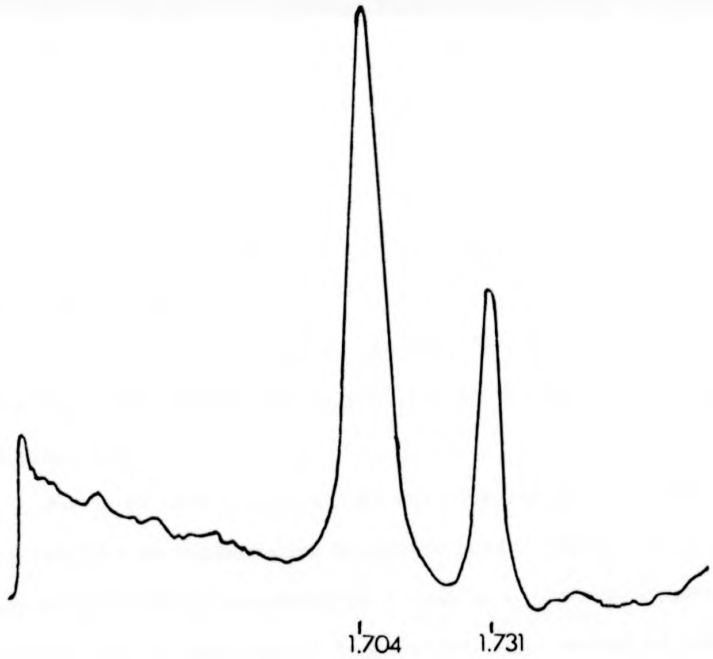
LEGEND TO FIG. 3.1

ANALYTICAL ULTRACENTRIFUGATION OF AXOLOTL DNA

5 μg native genomic axolotl DNA was mixed with 1 μg *Micrococcus lysodeikticus* DNA (ρ = 1.731) as marker in 0.1M TES. Ultracentrifuge grade CsCl was added to an initial density of 1.7 gm/cc. An aliquot (approximately 300 μl) was injected into the chamber of a Model E analytical ultracentrifuge. Centrifugation was for 22 hours at 42,000 rpm at 20°C. The gradient formed after 22 hours was photographed under u.v. light.

The negative was scanned using a Joyce-Loebl microdensitometer. Buoyant density of the axolotl DNA was calculated relative to the position of the marker DNA.


1.704
1.731


1.704
1.731

genetic differences in the organisation of the genome which are detectable by ultracentrifugation. Mammals and birds exhibit asymmetric banding patterns which are skewed towards the heavy (% G+C rich) side of the CsCl main band DNA. Amphibians, fish and reptiles tend to have more symmetrical distributions. Essentially symmetrical bands are shown by all invertebrates so far studied.

The data for the axolotl described above is similar to the data for the other amphibians examined by Thiery et al. These points will be discussed in more detail later.

The second method of analysis used to estimate the % G+C content was controlled thermal denaturation of genomic DNA. It has been shown by Mandel and Marmur (1971) that a relationship exists between the temperature at which 50% of a DNA sample becomes denatured (Tm) and the % G+C content. This can be given by

$$Tm = 69.3 + 0.41 (\% G+C)$$

which can be rewritten as

$$\% G+C = 2.44 (Tm - 69.3).$$

Thus the % G+C content can be obtained if the Tm is known (see also Appendix II).

The Tm of native axolotl DNA was measured by controlled heat denaturation in 0.12M sodium phosphate buffer pH6.8 in a Gilford spectrophotometer using a thermocuvette linked to a thermoprogrammer, allowing a constant rate of increase of temperature. The change in optical density relative to a blank cuvette was monitored constantly and printed directly onto a chart recorder. The hyperchromicity was taken as the change in optical density between 60°C and 100°C and the Tm was taken as the point at which 50% of the final hyperchromicity was reached.

The mean observed Tm was 88.5°C, the calculated % G+C was 46.8%. Figure 3.2a shows the melting curve replotted so that the hyperchromicity is given as a percentage. 3.2b shows the derivative plot of 3.2a.

LEGEND TO FIG. 3.2

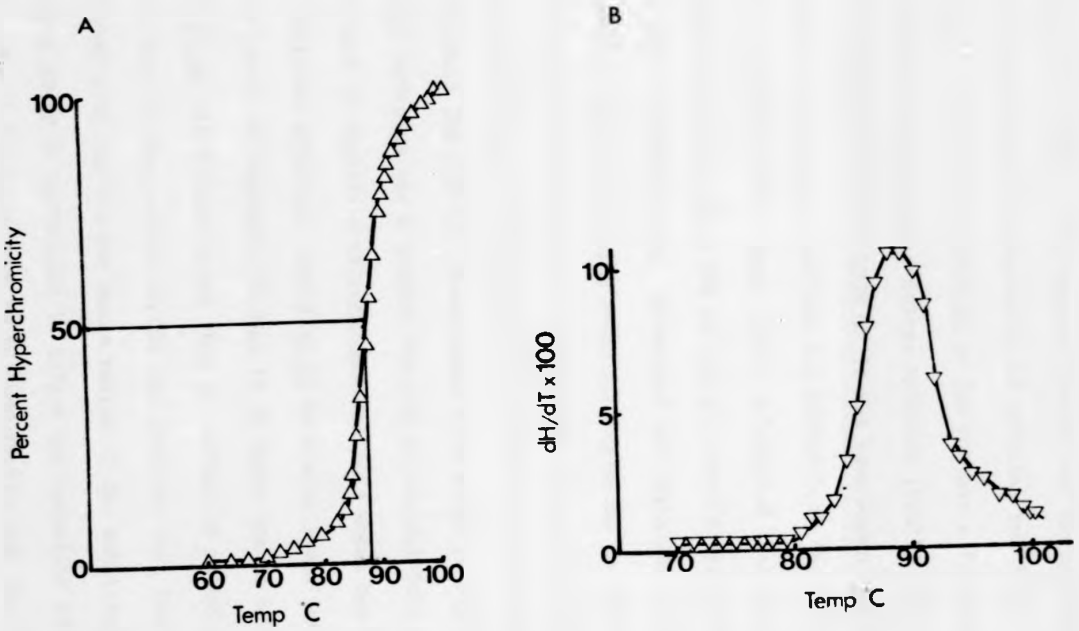
OPTICAL MELTING CURVE

Native genomic axolotl DNA was adjusted to an A_{260} of 1.0 in 0.12 M sodium phosphate buffer pH6.8. An aliquot was injected into the thermocuvette of a Gilford spectrophotometer linked to a Gilford Model 2527 Thermoprogrammer. 0.12M sodium phosphate was introduced into an adjacent chamber to act as a reference blank.

Temperature was increased at 0.5°C/min. and the change in absorbence was measured. Expansion effects were automatically corrected by reference to the blank sample.

Hyperchromicity was defined as the change in A_{260} between 60 and 98°C and the total change was recorded on a percentage scale (a).

The change in hyperchromicity per incremental increase in temperature is shown as the derivative plot (b).


A
100
50
0
Percent Hyperchromicity
60 70 80 90 100
Temp °C
B
10
5
0
dH/dT x 100
70 80 90 100
Temp °C

Several conclusions can be drawn from the combined results of the two sets of experiments. The % G+C content obtained from the two methods of analysis varies, from 45% from the buoyant density measurements to 46.8% from the Tm measurements. The average value would be 45.9%. The error may be due to differences in the experimental methods, or to other factors which may systematically affect both buoyant density and hyperchromicity measurements. One such factor is methylation of cytosine residues.

Most plants and animals studied so far contain a fraction of the cytosine residues in the form 5-methyl cytosine (Wyatt, 1951; Vanyushin *et al*, 1970; Kemp and Sutton, 1976). It has been known for some time that the methylation of cytosine residues may affect both the buoyant density and the Tm of a particular DNA. Kirk (1967) calculated that methylation of 25% of the cytosine residues in a DNA of 50% G+C would decrease the buoyant density in CsCl by 0.004 gm/cc. Szybalski and Szybalski (1971) have compared various DNAs containing methylated cytosine and they conclude that for each 1% methylation results in a decrease in buoyant density of 0.001gm/cc. David *et al* (1970) have investigated the effect of methylation on *Xenopus laevis* ribosomal DNA (rDNA). Chromosomal rDNA which is 67% G+C and has 45% 5-methyl cytosine has a buoyant density in neutral CsCl 0.0055 gm/cc less than that of amplified extrachromosomal rDNA which has the same % G+C but no methylated cytosine. Dawid *et al* have also shown that the effect on Tm of this level of 5-methyl cytosine is to raise the Tm by $3^{o}C$ in 0.18M Na^{+}.

Gill *et al* (1974) have noted that the effect of methylation on Tm is similar to that of bromination of DNA and conclude that the helix stabilisation is correlated with the Van der Waals radius of the substituting group. The substituting group is postulated to affect the hydration of the DNA, causing the DNA in the substituted form to be more stable than the non substituted form. The result would also tend to reduce the buoyant density of the DNA.

In contrast, Kemp and Sutton (1976), working with plant DNA, did not observe an increase in Tm with increasing methylation of DNA although they

noted that the buoyant densities of the DNAs were as predicted from the work of Kirk (1967, noted above). Recently Wagner and Capesius (1981) have suggested that although buoyant density and methylation are connected the Tm is increased only when 5-methyl cytosine represents 30% or more of all of the cytosine residues.

It has been known for some time that amphibian genomic DNA contains methylated cytosine. Dawid et al (1970) observed that 1.3% of total Xenopus laevis DNA was 5-methyl cytosine, while Vanyushin et al (1970) have reported a value of 1.6% for Rana temporaria. Bird and Taggart (1980) have shown by restriction enzyme analysis that DNA from both Xenopus and Triturus cristatus is not restricted by enzymes which do not cut at methylated sites (eg. Hpa II) compared to isoschizomes which cut at methylated sites (eg. Msp I, see Waalwijk and Flavell, 1978).

Based on the small differences between the % G+C values obtained from the buoyant density and Tm data, and considering the published data on 5-methyl cytosine content in amphibia it is not expected that the level of methylation would be greatly different from that of X. laevis, indeed methylation in the order of 1 - 1.5% of bases would account for the observed data based on the considerations noted above.

The presence of 5-methyl cytosine was qualitatively demonstrated by the use of the restriction endonucleases Hpa II and Msp I. Hpa II recognises and hydrolyses within the sequence 5'-CCGG but not 5'CMeCGG whereas Msp I hydrolyses DNA within both sequences. The results of one such experiment are shown in Fig. 3.3A, and a densitometric scan of the tracks is shown in Fig. 3.3B.

It is clear that the axolotl DNA is more resistant to the enzyme Hpa II than to Msp I. In this experiment digestion conditions were such that an equivalent amount of PBR322 DNA was completely digested to give the same restriction pattern for each enzyme (see Sutcliffe, 1978). However, it is possible that some factor in the axolotl DNA preparation

LEGEND TO FIG. 3.3

QUALITATIVE DEMONSTRATION OF METHYLATION AT THE SEQUENCE 5'CCGG

2 μg of axolotl DNA was digested to completion by overnight incubation at 37°C with 5 units of the relevant restriction endonuclease in a volume of 20 μl in the buffer suggested by the supplier. The reaction was terminated by the addition of 1/10th volume of 100mM Tris pH7.5; 100mM EDTA; 0.3% agarose beads; 0.01% orange G; 50% glycerol. The samples were electrophoresed in a 1.5% agarose gel containing 1 μg/ml ethidium bromide for 16 hours at 30 V. The gel was illuminated with u.v. light (360 nm) and photographed.

Track a. Msp-I

b. HpaII

c. Hae-III

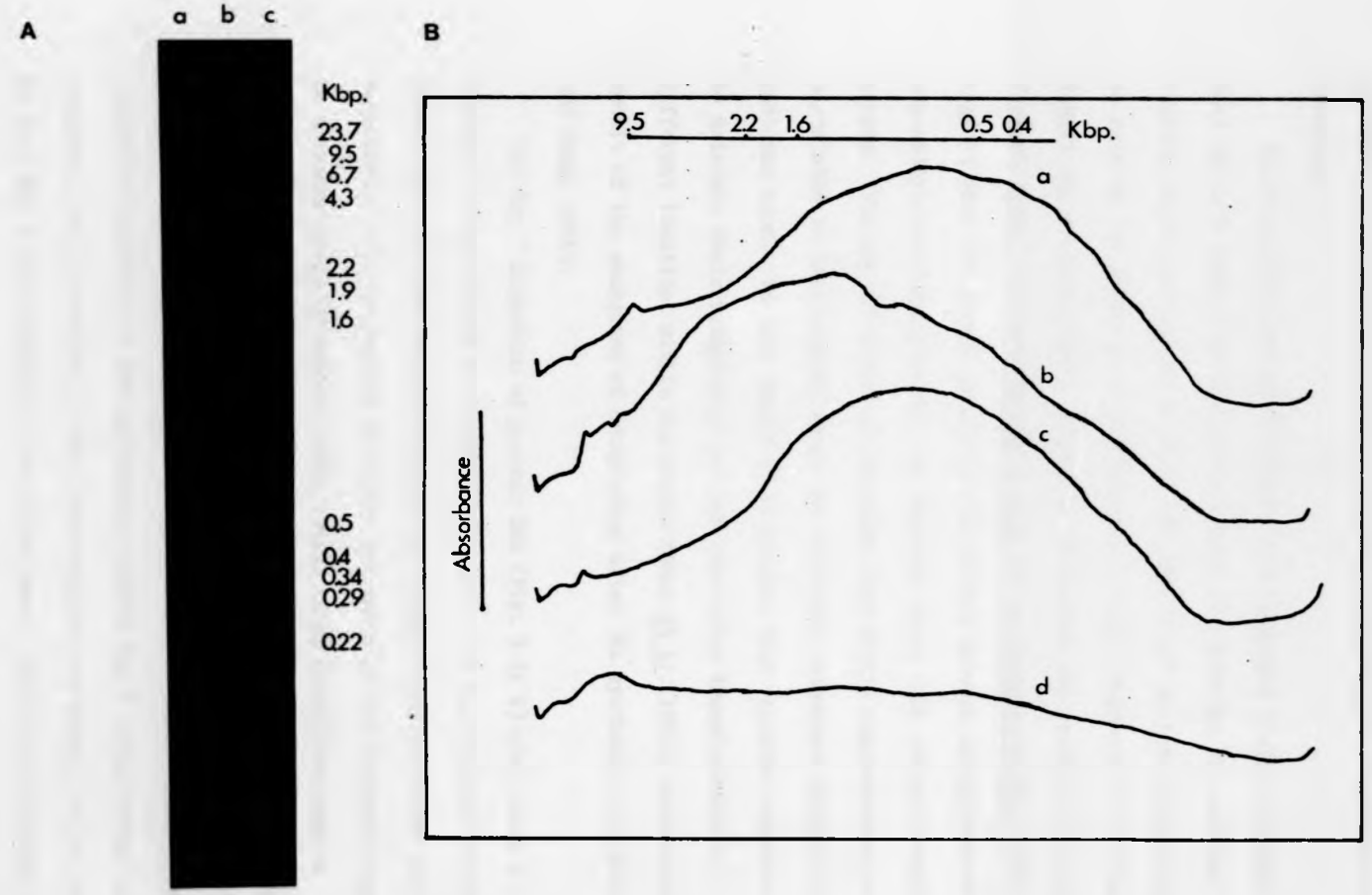
The negative was scanned using a Joyce-Loebl microdensitometer

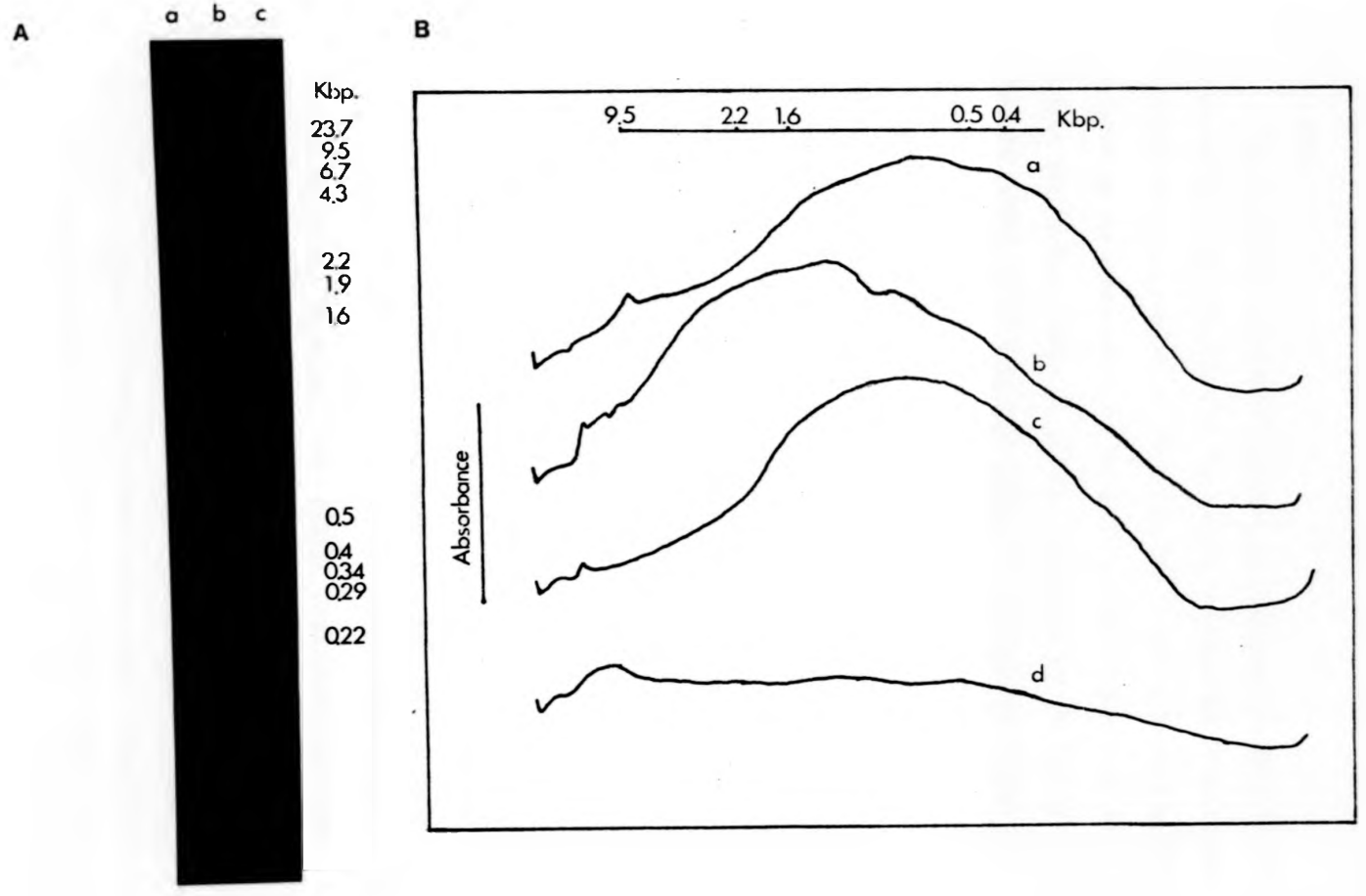
a-c as above

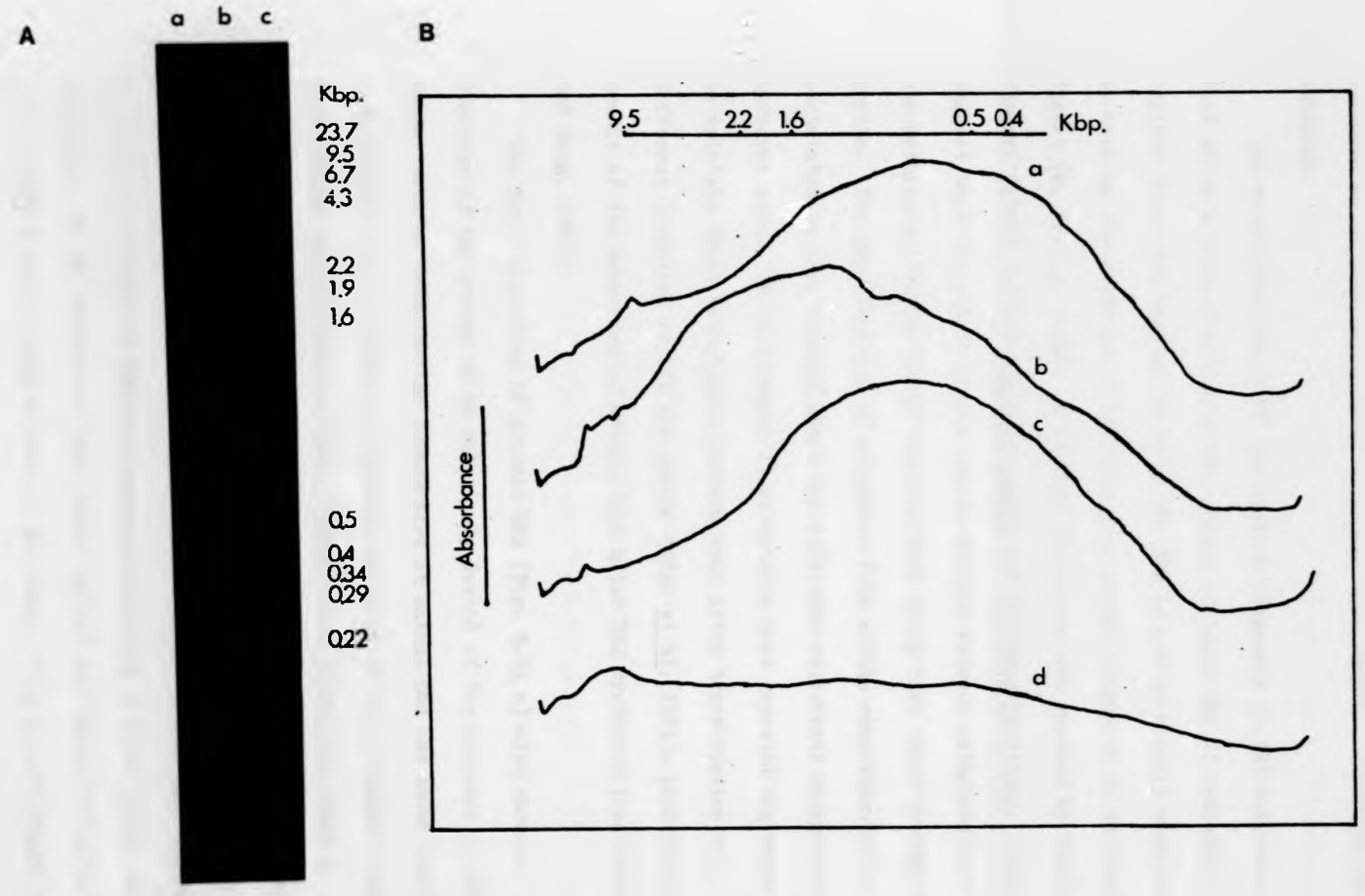
d. background fluorescence

Molecular weight markers were HindIII digested λ DNA

and HinfI digested pBR322 DNA.


A
a b c
Kbp.
23.7
9.5
6.7
4.3
2.2
1.9
1.6
0.5
0.4
0.34
0.29
0.22
B
9.5
2.2
1.6
0.5 0.4
Kbp.
a
b
c
d
Absorbance


A
a b c
Kbp.
23.7
9.5
6.7
4.3
2.2
1.9
1.6
0.5
0.4
0.34
0.29
0.22
B
9.5
2.2
1.6
0.5 0.4
Kbp.
a
b
c
d
Absorbance


A
a b c
Kbp.
23.7
9.5
6.7
4.3
2.2
1.9
1.6
0.5
0.4
0.34
0.29
0.22
B
9.5
2.2
1.6
0.5 0.4
Kbp.
a
b
c
d
Absorbance

was preventing total hydrolysis. In later experiments, which gave the same result (see Figs. 5.13 - 5.15) a small amount of phage λ DNA was included as an internal control to ensure that complete digestion had occurred.

The size distribution of restriction fragments (3.3B) indicates that while a large fraction of the genome contains Hpa II resistant regions there are regions in which the DNA is not as heavily methylated, as seen by the presence of low molecular weight fragments in the Hpa II digest (b, in Fig. 3.4B). A similar conclusion was reached by Bird and Taggart (1980) for both Xenopus laevis and Triturus cristatus. They suggest that the animal genome can be divided between methylated and non-methylated "compartments" interspersed among each other through the genome. The amplification of sequences from within compartments is postulated to have brought about the different vertebrate methylation patterns seen. In this regard it is notable that repeated sequences tend to maintain their methylation pattern even after translocation to different locations within the genome (Eden et al, 1981), probably as a result of the mechanism of methylation after DNA synthesis (see Ehrlich and Wang, 1981).

The Msp I digestion of genomic DNA (Fig. 3.3A a) also shows a fraction of the genome to be relatively devoid of the sequence 5'-CCGG. Recent data on dinucleotide frequencies in animal DNA has shown that the dinucleotide CpG is present at around one third of the frequency expected on a random basis (Nussinov, 1980, 1981). Bird (1980) has shown a correlation between low CpG levels and high levels of methylation and suggests that 5MeCpG tends to mutate to TpG. Indeed TpG is the most frequently observed dinucleotide (Nussinov, 1980, 1981). Thus the Msp I resistant stretches of DNA may contain mutated Msp I sites within the sequence. An alternative is that these regions are simply of low %G+C so that Msp I sites would be rare in any case. This is not thought to

was preventing total hydrolysis. In later experiments, which gave the same result (see Figs. 5.13 - 5.15) a small amount of phage λ DNA was included as an internal control to ensure that complete digestion had occurred.

The size distribution of restriction fragments (3.3B) indicates that while a large fraction of the genome contains Hpa II resistant regions there are regions in which the DNA is not as heavily methylated, as seen by the presence of low molecular weight fragments in the Hpa II digest (b, in Fig. 3.4B). A similar conclusion was reached by Bird and Taggart (1980) for both Xenopus laevis and Triturus cristatus. They suggest that the animal genome can be divided between methylated and non-methylated "compartments" interspersed among each other through the genome. The amplification of sequences from within compartments is postulated to have brought about the different vertebrate methylation patterns seen. In this regard it is notable that repeated sequences tend to maintain their methylation pattern even after translocation to different locations within the genome (Eden et al, 1981), probably as a result of the mechanism of methylation after DNA synthesis (see Ehrlich and Wang, 1981).

The Msp I digestion of genomic DNA (Fig. 3.3A a) also shows a fraction of the genome to be relatively devoid of the sequence 5'-CCGG. Recent data on dinucleotide frequencies in animal DNA has shown that the dinucleotide CpG is present at around one third of the frequency expected on a random basis (Nussinov, 1980, 1981). Bird (1980) has shown a correlation between low CpG levels and high levels of methylation and suggests that 5MeCpG tends to mutate to TpG. Indeed TpG is the most frequently observed dinucleotide (Nussinov, 1980, 1981). Thus the Msp I resistant stretches of DNA may contain mutated Msp I sites within the sequence. An alternative is that these regions are simply of low %G+C so that Msp I sites would be rare in any case. This is not thought to

be the case. Hae III, which cuts within the sequence 5'GGCC, the reverse of the Msp I sequence but containing two of the three possible dinucleotide pairs found in the Msp I sequence does not show these high molecular weight fragments when used to digest axolotl DNA (Fig. 3.4A, c).

Table 3.2 shows some data on % G+C content in animals collected from various sources. Some of the data has been derived from buoyant density measurements and some from thermal denaturation data, the different methods used to obtain the data are noted, and the significance of the differences in result obtained by the two methods is discussed later.

From the table it can be seen that a spread of % G+C value exists with a modal value around 42%. Species with low C values appear to be more A + T rich (insects, echinoderms), however sufficient exceptions exist so that little can be definitely concluded as to whether the C value and % G+C content are correlated. It may be however that a general increase in % G+C content with increasing C value may be a factor in reducing the skewed distributions noted by Thiery et al (1976). It is of interest that of the Urodeles which have been studied, all have % G+C contents of 45% or more compared to the value of 40.9% for X. laevis an Anuran. Whether this represents a true reflection of the phylogenetic relationship remains to be seen.

From the melting profile (Fig. 3.2b) it can be seen that a small fraction of the DNA appears to melt at a higher temperature than the main body of DNA. This may be due, as has been suggested above, to local regions of heavily methylated DNA which may affect the Tm of a fraction of the genome. Alternatively a GC rich satellite fraction, undetected in the analytical centrifugations may exist. The possible existence of such a satellite was investigated.

Walker defined satellite DNA as a native fraction of the chromosomal DNA which after isolation by any method gives a narrow unimodal band in

TABLE 3.2

%G+C CONTENT IN ANIMALS

		%G+C
MAMMALS		
	HOMO SAPIENS	41.8[a]
	BOS TAURUS	44.5[a]
	FELIS DOMESTICUS	44[a]
	CANIS FAMILIARIS	45[a]
	CAVIA PORCELLUS	41.8[a]
	MUS MUSCULUS	41.6[a]
	ORYCTOLAGUS CUNICULUS	44.2[a]
AMPHIBIA		
	XENOPUS LAEVIS	40.5[a]
	PLEURODELES WALTII	45.6[a]
	PLETHODON CINEREUS	45[b]
	NOTOPTHALMUS VIRIDESCENS	46[c]
	AMBYSTOMA MEXICANUM	45[d]
BIRDS		
	GALLUS DOMESTICUS	44, 40.8*[e]
	Pigeon	38.8*[e]
	Duck	40.8*[e]
REPTILES		
	IGUANA IGUANA	43.1[a]
	Turtle	41.7*[f]
	Python	41.3*[f]
	Crimon	42.7*[f]
FISH		
	SALMO SALAR	44.4[a]
	SALMO IRIDEUS	44.1*[g]
	SPRATTUS SPRATTUS	46.3*[g]
	CLUPEA HARENGUS	47.2*[g]
	CYPRINUS CARPIO	40.2*[g]
	CALLIPHERA VICINA	29.6, 31

TABLE 3.2

%G+C CONTENT IN ANIMALS

	%G+C
MAMMALS	
HOMO SAPIENS	41.8[a]
BOS TAURUS	44.5[a]
FELIS DOMESTICUS	44[a]
CANIS FAMILIARIS	45[a]
CAVIA PORCELLUS	41.8[a]
MUS MUSCULUS	41.6[a]
ORYCTOLAGUS CUNICULUS	44.2[a]
AMPHIBIA	
XENOPUS LAEVIS	40.5[a]
PLEURODELES WALTII	45.6[a]
PLETHODON CINEREUS	45[b]
NOTOPTHALMUS VIRIDESCENS	46[c]
AMBYSTOMA MEXICANUM	45[d]
BIRDS	
GALLUS DOMESTICUS	44, 40.8*[e]
Pigeon	38.8*[e]
Duck	40.8*[e]
REPTILES	
IGUANA IGUANA	43.1[a]
Turtle	41.7*[f]
Python	41.3*[f]
Crimon	42.7*[f]
FISH	
SALMO SALAR	44.4[a]
SALMO IRIDEUS	44.1*[g]
SPRATTUS SPRATTUS	46.3*[g]
CLUPEA HARENGUS	47.2*[g]
CYPRINUS CARPIO	40.2*[g]
CALLIPHERA VICINA	29.6, 31

TABLE 3.2 (cont.)

	%G+C
INSECTS	
DROSOPHILA MELANOGASTER	42.2[a]
MUSCA DOMESTICA	37.5*[h]
CHIRONOMUS TENTANS	33*[j]
CHIRONOMUS THUMMI	28.6, 28.5*[k]
CALIPHERA VICINA	29.6, 31*[l]
ECHINODERMS	
STRONGYLOCENTROTUS PURPURATUS	40.4[a]

%G+C calculated from : $\%G+C = (\rho - 1.660)/0.00098$

or : $\%G+C = (Tm - 69.3)/0.41$

Data collected from buoyant density measurements except where noted (*)

REFERENCES

a. Thiery et al (1976)

b. MacGregor and Kezer (1971)

c. Barsacchi and Gall (1972)

d. This thesis

e. Epplen et al (1978)

f. Epplen et al (1979)

g. Hanham and Smith (1979)

h. Crain et al (1976)

j. Wells et al (1976)

k. Wobus (1975)

l. Scheller and Mashenk (1978)

CsCl because of common properties shared by its sequence (Walker 1971). Satellite DNAs were defined as such specifically to exclude bands obtained simply by fractionating a preparative CsCl gradient, low molecular weight fractions which will not band simply in CsCl, and fractions produced by reassociation of native DNA which give very sharp bands due to high molecular weight networks. By this criterion the satellites of Xenopus laevis, Pleurodeles waltii (Thiery et al, 1976) and Triturus cristatus carnifex (Varley et al, 1980 b) would not be recognised as such. Such satellites as these can be isolated by binding metal ions eg. Ag^{+}, Hg^{2+} to the DNA and centrifugation in Cs_2SO_4, a more dense salt than CsCl. Although it is known that Ag^{+} binds to GC rich DNA whereas Hg^{2+} binds to AT rich DNA the banding profile in such gradients can be complex and it is obvious that the GC content is not the only factor involved in determining the position of satellites in such gradients. Nevertheless fractions can be obtained which satisfy Walker's criterion. Further such fractions can usually be shown to exhibit those other characteristics attributed to classical "satellite" DNA i.e. to be located primarily in heterochromatic regions; to have a simple repetitive sequence and to be organised as long tandem arrays (Drosophila satellites can be separated from main band DNA at fragment lengths of up to 800 kbp (Brutlag et al, 1977).

In view of these considerations it was decided to examine the axolotl genome in some detail for the presence of a satellite fraction. Livesey (1980) has shown that axolotl chromosomes can be stained by the C banding technique of Pardue and Gall (1970). It is widely accepted that there is a correlation between C banding centromeric heterochromation and the presence of satellite DNA, although the relationship is not absolute. Therefore preparative CsCl centrifugations of native axolotl DNA were performed in order to determine whether any satellite could be present that had been overlooked in the analytical centrifugations. One

such gradient is shown in Fig. 3.4. No obvious satellite can be seen in these preparative gradients.

When fractions containing the heavy side of the gradient were pooled from several parallel gradients and recentrifuged they still showed a unimodal distribution indicating only one major buoyant density class. From these experiments it would appear that no major satellite fraction exists. A similar conclusion has been reached by workers at the Axolotl Colony University of Indiana (Malacinski; pers. comm.). However as noted earlier, both Thiery et al (1976) and Varley et al (1980b) showed that satellite like fractions could be isolated by Cs_2So_4/Ag^+ gradient centrifugation. Several gradients were prepared by the methods of Varley et al (1980b), varying the Ag/DNA-P ratio from 0.275 to 0.35, this being the range commonly used to isolate satellite fractions. Again no obvious satellite fraction was obtained (data not shown). This does not mean that tandem arrays of similar repeated sequences do not exist within the axolotl genome, but does suggest that such arrays are not as extensive as in other amphibians where such studies have been performed.

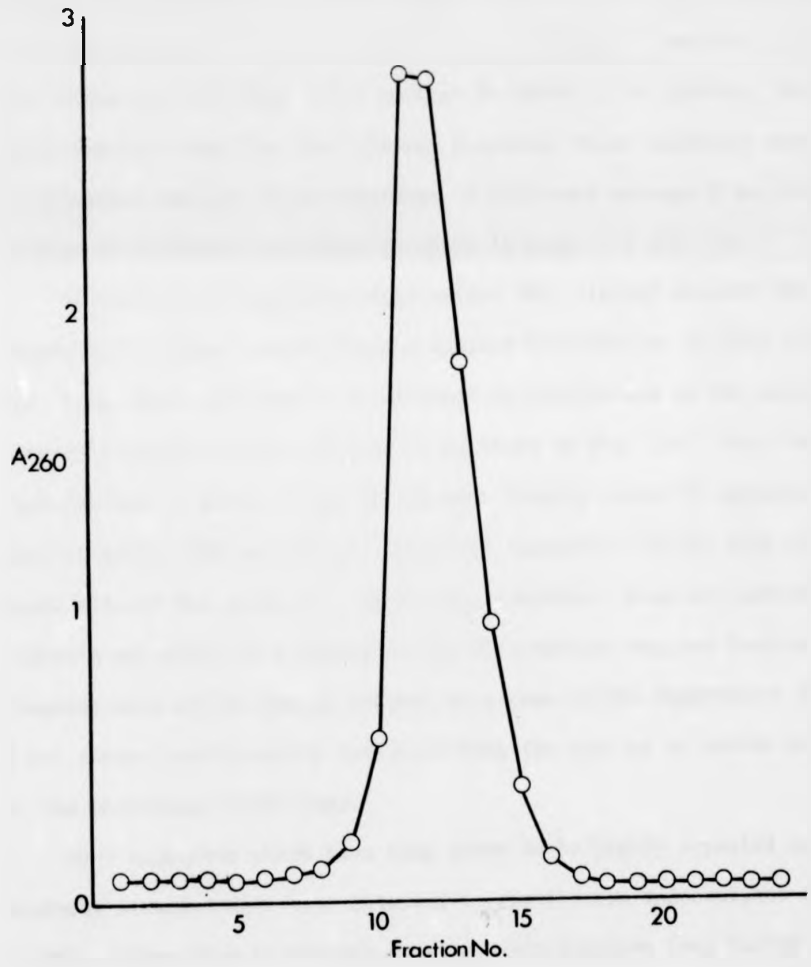
It therefore appears that while an association between satellite DNA and centromeric C banding regions has been found in several instances, in the case of the axolotl this does not hold. It is possible that there may be a requirement for a particular sequence organisation which will give rise to C bands under appropriate conditions, and that in certain cases the sequences involved become amplified into long tandem arrays which can be isolated as satellite DNA. It is of interest that the recently sequenced centromeric DNA of Saccharomyces cerevisiae contains sequences with significant homology to various satellite DNA sequences from higher eukaryotes (Fitzgerald-Hayes et al, 1982). Thus different organisms may have common sequences present in variable copy number within centromeres.

In order to determine whether repetitive sequences with extremes of GC content do exist in the axolotl the following experiment was performed.

LEGEND TO FIG. 3.4

PREPARATIVE CsCl CENTRIFUGATION OF AXOLOTL DNA

Axolotl DNA was mixed with a saturated solution of CsCl to an initial buoyant density of 1.710 gm/cc. The gradients were centrifuged at 40,000 rpm, 40 hours at 25°C. The gradients were fractionated by puncturing the bottom of the tube and collecting fractions. The A_{260} of each fraction was measured after addition of H_2O to 400 µl total volume per fraction.


3
2
1
0
A260
5
10
15
20
Fraction No.

High molecular weight axolotl DNA was denatured and reassociated to an observed Cot of 1. Single stranded DNA was removed by S_1 nuclease and the duplex DNA was labelled by "nick translation" for a short period of time. The labelled DNA was cocentrifuged with 50 µg. of cold, high molecular weight DNA in a CsCl gradient. The results of one such gradient is shown in Fig. 3.5. The observed range of buoyant densities indicates that sequences with high % G+C content do exist in the genome. However taken together with the data already presented these sequences must be interspersed amongst other sequences of different average % G+C content to give the unimodal distribution shown in Figs. 3.1 and 3.4.

In contrast to high molecular weight DNA, lightly sheared DNA (average length 2.5 kilobase pairs) shows a similar distribution to that shown in Fig. 3.5b. When such DNA is centrifuged to equilibrium in the analytical ultracentrifuge the distribution is as shown in Fig. 3.6. Here two major features can be seen. First the buoyant density spread is greater than that of native DNA and second a distinct 'shoulder' can be seen on the heavy side of the gradient. While this 'shoulder' does not satisfy Walker's criterion of a satellite the observations suggest that as the fragment size of the DNA is reduced stretches of DNA appear which have a % G+C content sufficiently different from the mean as to become noticeable in the analytical centrifuge.

Many sequences which have been shown to be highly repeated and tandemly arranged have been detected by restriction endonuclease digestion of DNA, rather than by buoyant density centrifugation (see Varley et al, 1980b for example). Therefore axolotl DNA was cleaved with a variety of enzymes recognising four, five or six base pair sequences. Several digests produced bands indicating sequences with a common unit length bounded by two enzyme cleavage sites, however in all but one case examined the pattern characteristic of a tandem repeat could not be seen. The exception was Hae III which produced a major band of approximately 60 base pairs and an

LEGEND TO FIG. 3.5

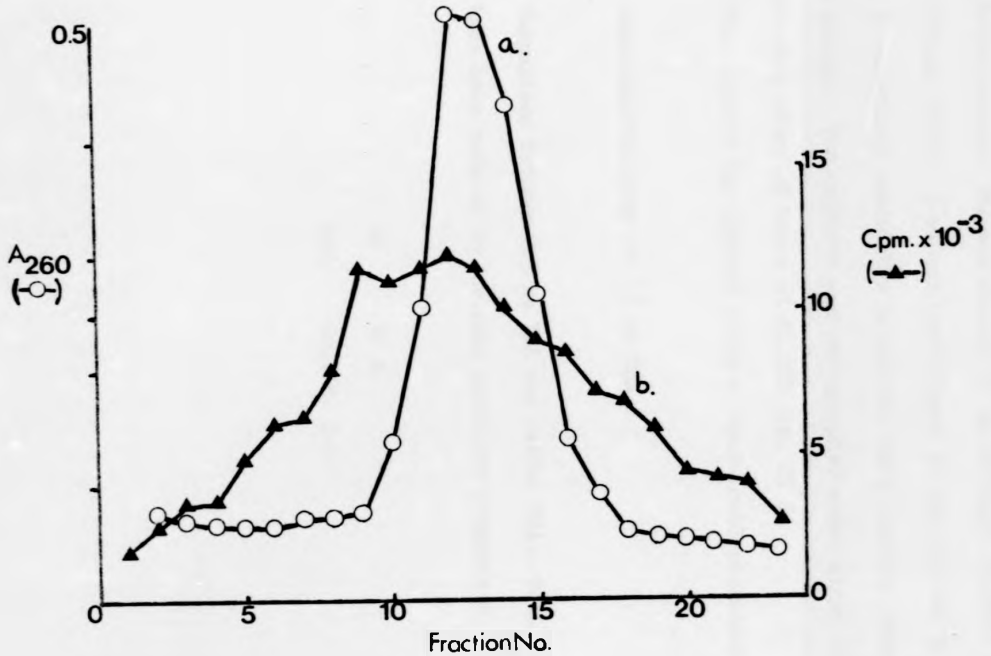
G+C DISTRIBUTION OF HIGHLY REPETITIVE AXOLOTL DNA

Native axolotl DNA was denatured and reassociated to an observed Cot of 1. The DNA was digested with an excess of S_1 nuclease, precipitated, redissolved and 1 μg was 'nick translated' for a short time, without DNAase I, using ^{32}P dCTP. The labelled DNA was passed over Sephadex G-50 and the exclusion peak was pooled. 50 μg of genomic DNA and approximately 10^5 Cerenkov cpm of ^{32}P repetitive Cot fraction were mixed.

CsCl gradients were prepared to give an initial buoyant density of 1.710 gm/cc and centrifuged and processed as Fig. 3.5. Each fraction was counted by Cerenkov counting before the A_{260} was measured.

a. (- O -) Absorbence at 260 nm

b. (- ▲ -) Cerenkov cpm.


0.5
a.
15
A_{260}
(-○-)
Cpm. x 10^{-3}
(-▲-)
10
b.
5
0
0
5
10
15
20
Fraction No.

LEGEND TO FIG. 3.6

ANALYTICAL ULTRACENTRIFUGATION OF 2.5kbp AXOLOTL DNA

Native axolotl DNA was sheared to an average fragment length of 2.5 kilobase pairs. 5 µg was centrifuged to equilibrium in a Beckman Model E analytical centrifuge in neutral CsCl (initial buoyant density 1.710 gm/cc). The gradient was photographed under ultra violet illumination after 22 hours at 42,000 rpm, 25°C.

The negative was scanned using a Joyce-Loebl microdensitometer

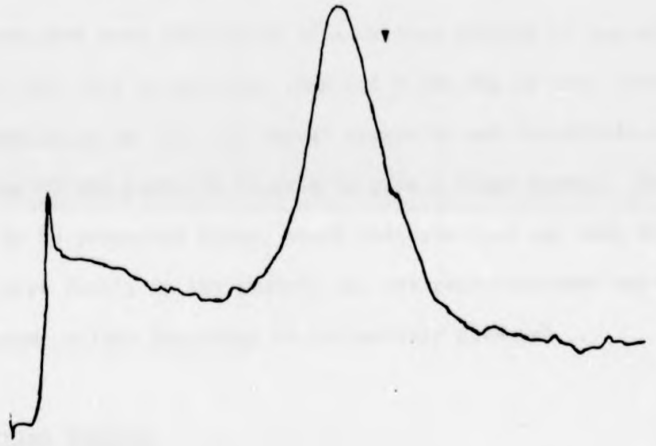
a. Densitometric scan of 2.5 kb DNA

b. Comparison between 2.5 kb DNA and native DNA. Both figures have been reduced to the same scale for comparison.

bi as a.

bii as Fig. 3.1

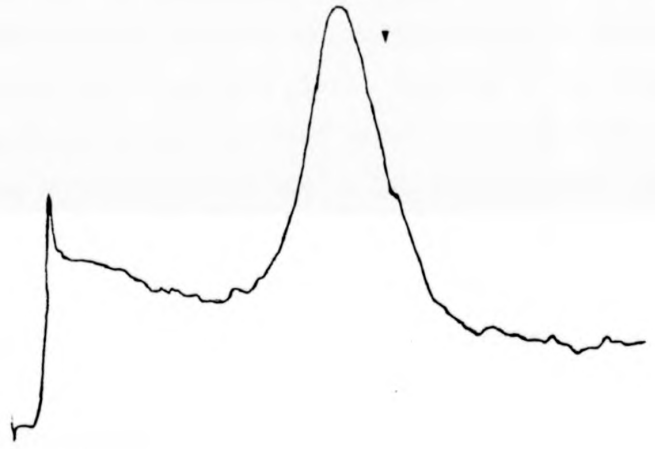
a.



b.i

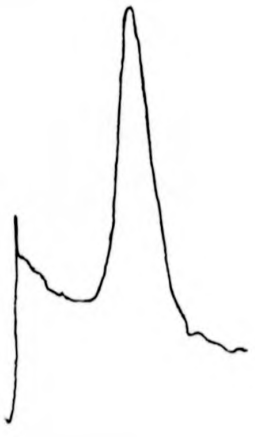
b.ii

a.



b.i

b.ii



indication of multiples of the monomer which can be produced from a tandem array by random inactivation of sites by mutation. Fig. 3.7 shows an Hae III digest of axolotl DNA. The 60 bp (approx.) band is clearly visible. As will be shown later this band is arranged as tandem arrays within the genome. However when Cot 0-50 or 0-100 duplex DNA was used to probe Hae III digested axolotl DNA no clear picture emerged. Cot 0-50 DNA does give some indication of a banding pattern at low molecular weights but this is not seen when Cot 0-100 DNA is used (data not shown). One conclusion is that the repeat arrays do not constitute a large enough fraction of the probe to be able to give a clear signal. The reassociation data, to be presented later, would indicate that any such highly repetitive family in the axolotl can not represent much more than 1% of the genome unless the array is extensively diverged.

CONCLUDING REMARKS

From this short examination of various physical parameters of the axolotl genome several features of the genome have been identified. The haploid DNA content (C value) is 38 pg. The % G+C content is between 45 and 46.8%, based on a Tm in 0.18M Na^+ of 88.5°C and a buoyant density in neutral CsCl of 1.704 g/cc. The genome contains 5-methyl cytosine at an estimated level of 1 - 1.5% of bases. No satellite DNA has been identified in neutral CsCl or $CsSo_4/Ag^+$ gradients.

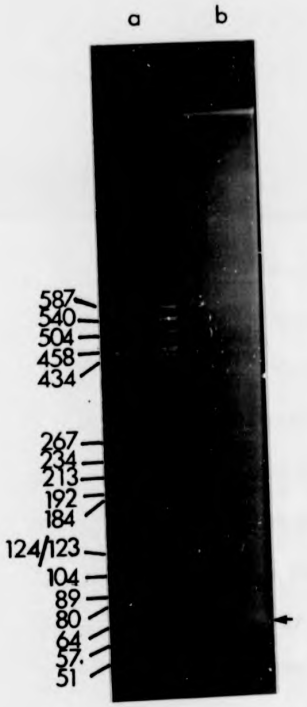
LEGEND TO FIG. 3.7

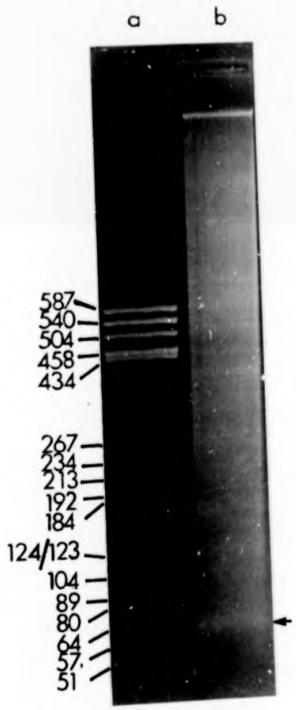
HaeIII DIGESTION OF AXOLOTL DNA

2 μg axolotl genomic DNA was digested to completion with HaeIII and the fragments were separated by size on 2.5% agarose gels.

a. pBR322 - HaeIII. Fragment lengths are from Sutcliffe, (1978.)

b. Axolotl - HaeIII.


a
b
587
540
504
458
434
267
234
213
192
184
124/123
104
89
80
64
57
51


a
b
587
540
504
458
434
267
234
213
192
184
124/123
104
89
80
64
57
51

RESULTS AND DISCUSSION II

REASSOCIATION KINETIC ANALYSIS OF AXOLOTL REPETITIVE DNA

In this section the axolotl genome will be analysed by reassociation kinetics to demonstrate the presence of repetitive sequences within the genome and to characterise these sequences in terms of repetition frequency, interspersion pattern within the genome and evolutionary divergence.

1. Reassociation kinetics

Appendix II discussed the relevant theory of reassociation kinetics. For a simple DNA, i.e. a genome in which analytical complexity is equivalent to sequence complexity, the rate of reassociation is proportional to the square root of the fragment length (Wetmur & Davidson, 1968). For a complex genome containing interspersed repeated sequences the observed rate of reassociation, when analysed by hydroxylapatite chromatography, is determined by the most highly repetitive element present on each fragment. It is therefore necessary to shear the DNA such that on average each fragment contains only one repetition frequency class. In practice it has been shown that a fragment size of around 3 - 500 nucleotides satisfies these conditions.

Several methods exist for shearing DNA including sonication, high pressure, ejection from a French Press, high speed homogenisation or enzymic digestion by deoxyribonucleases. Several of these methods were tried to determine the most practicable.

Treatment of axolotl DNA with varying concentrations of DNAase I for different lengths of time lead to a wide spread of fragment sizes from oligomers to fragments of several kilobases in length. The available French Press was reported to yield fragments of around 1000 nucleotides (Potts, 1981; Avery, pers. comm.). Passage through a G.25 needle using

maximum thumb pressure produced only large single strand fragment lengths after up to 20 passages. Sonication, on ice, was found to produce fragments with an average single strand size of around 500 bases after 10 x 30 seconds bursts with 30 seconds cooling between bursts, using a Dawes soniprobe, setting No. 1 tuned to 1mA. This method was used to prepare all of the short fragments used in the reassociation experiments. After sonication the DNA was passed over a Chelex 100 exchange column to remove any contaminating heavy metal ions before precipitation (see Appendix II : Affects on Rate).

It was observed that the fragment size varied in some instances between batches of DNA. This may be due to such factors as the positioning of the soniprobe tip, the final volume of solution, the shape of the vessel containing the DNA solution. The best practice was to sonicate a large batch of DNA at one time, determine the single strand size and store at -20°C until required.

Figure 4.1 shows the reassociation of axolotl DNA and E. coli DNA over the range of Cot from 10^{-5} to 10^{5}. To achieve the extremes of Cot, both high and low, various initial DNA concentrations were employed, along with several salt concentrations. All of the data points have been corrected to the Cot equivalent to reassociation at 60°C in 0.12M sodium phosphate (NaPB) buffer pH6.8 (see Britten et al, 1973 for the salt conversion table). The axolotl DNA has an observed Tm of 88.5% in 0.12M NaPB (Section 1). Therefore reassociation at 60°C in 0.12M NaPB (or equivalent) occurs at Tm -28.5. This is within the optimal range of temperatures for maximal rate of reassociation (Bonner et al, 1973) and allows hybrids with up to 28% mismatching to occur (Bonner et al, 1973). However it should be noted that hybrids with Tm at or near the temperature of reassociation may not form and so be under represented in subsequent analysis.

The smooth line through the data points was obtained by analysing

LEGEND TO FIG. 4.1

REASSOCIATION KINETICS OF SHEARED AXOLOTL & E.coli DNA

DNA was sheared by sonication to an approximate weight average single strand length of 500 nucleotides. The DNA was passed over Chelex 100, precipitated and redissolved at concentrations ranging from 0.05 to 5 mg/ml in 0.12M NaPB pH6.8. The DNA was denatured by boiling for 5 minutes at 100°C. The salt concentration was adjusted as required and the samples were overlaid with paraffin oil and incubated at 60°C or the equivalent temperature required to maintain the same criterion.

At various times 50 µg aliquots were removed and made 1 ml and 0.12M NaPB using ice cold solutions to quench the reactions. The samples were quick frozen in dry ice acetone and stored at -20°C for as short a time as possible before fractionation. The samples were fractionated by hydroxylapatite chromatography. Data was plotted as % totally single stranded vs log equivalent Cot (Log E Cot) (Britten et al, 1973).

A non linear least squares solution to the data was fitted using the computer program of Kells and Straus (1977), modified by myself and Mr. N. Coveney of the University of Warwick Computer Unit, for use on a Burroughs B6700 computer.

Axolotl genomic DNA (●)

E.coli DNA (o)

LEGEND TO FIG. 4.1

REASSOCIATION KINETICS OF SHEARED AXOLOTL & E.coli DNA

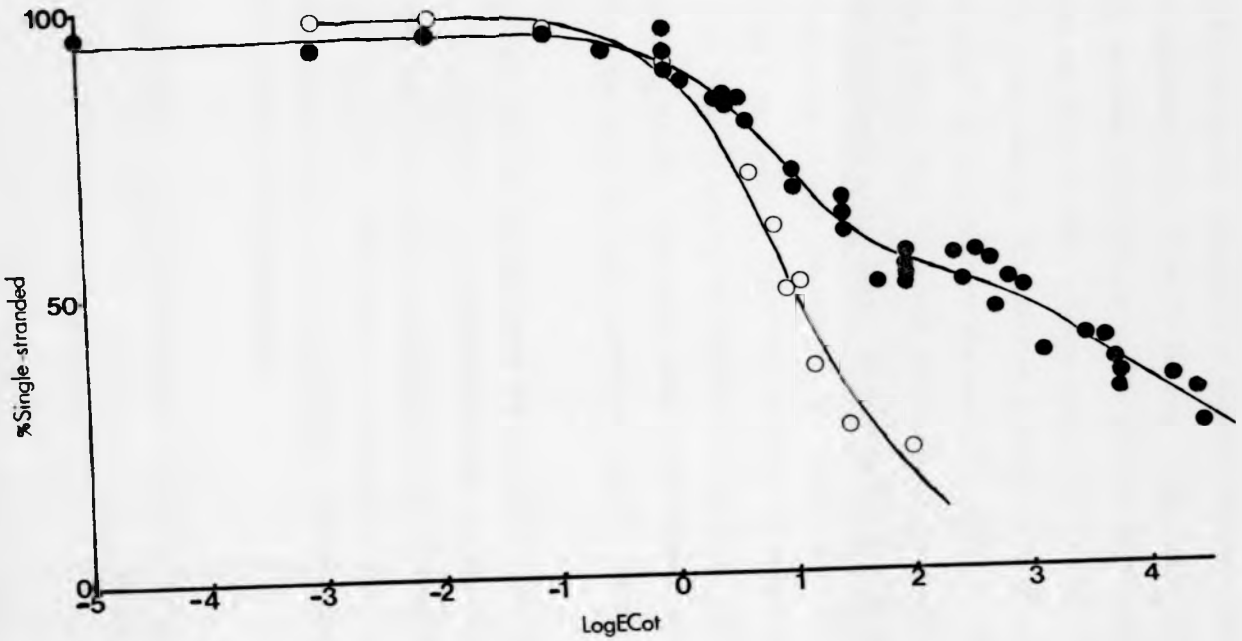
DNA was sheared by sonication to an approximate weight average single strand length of 500 nucleotides. The DNA was passed over Chelex 100, precipitated and redissolved at concentrations ranging from 0.05 to 5 mg/ml in 0.12M NaPB pH6.8. The DNA was denatured by boiling for 5 minutes at 100°C. The salt concentration was adjusted as required and the samples were overlaid with paraffin oil and incubated at 60°C or the equivalent temperature required to maintain the same criterion.

At various times 50 µg aliquots were removed and made 1 ml and 0.12M NaPB using ice cold solutions to quench the reactions. The samples were quick frozen in dry ice acetone and stored at -20°C for as short a time as possible before fractionation. The samples were fractionated by hydroxylapatite chromatography. Data was plotted as % totally single stranded vs log equivalent Cot (Log E Cot) (Britten et al, 1973).

A non linear least squares solution to the data was fitted using the computer program of Kells and Straus (1977), modified by myself and Mr. N. Coveney of the University of Warwick Computer Unit, for use on a Burroughs B6700 computer.

Axolotl genomic DNA (●)

E.coli DNA (o)


100
50
0
%Single-stranded
-5
-4
-3
-2
-1
0
1
2
3
4
LogECot

the data using a non linear least squares regression employing the computer programme of Kells and Straus (1977). The programme was modified for use on the University of Warwick Burroughs B6700 computer with the invaluable help of Mr. N. Coveney of the Computer Unit. The programme was used in preference to that given in Britten et al (1973) because it can reach a final solution using fewer iterative steps (D. Kohne, pers. comm.). The modified program is listed in Appendix II.

The solutions obtained from the curves are given in Table 4.1. It should be noted at this stage that resolution of the genome into a small number of components is illusory. It represents the simplest solution that can be drawn through the points. The values given for the components noted Class I and Class II repetitive in Table 4.1 represent averages for a whole range of repetitive sequence families with varying copy number and hence $Cot_{\frac{1}{2}}$ values. This point needs to be stressed to avoid the assumption that only two major repetitive sequence classes exist in the axolotl.

The $Cot_{\frac{1}{2}}$ pure for E. coli under the conditions used was found to be 8. This is comparable to published values and has been used as a standard for single copy reassociation, assuming an analytical complexity of 4.5×10^6 base pairs (Straus, 1971). The $Cot_{\frac{1}{2}}$ for unique sequences, if all of the axolotl genome were to reassociate with single copy kinetics at this criterion, would be given by

$$Cot_{\frac{1}{2}}(\text{axolotl}) = \text{Analytical complexity (axolotl)} \times \frac{Cot_{\frac{1}{2}}\ (\text{E. coli})}{\text{Analytical complexity (E. coli)}}$$

Axolotl DNA has an observed genome size of 37.7 pg. This corresponds to 3.64×10^{10} base pairs for a genome of 46% G+C. Therefore if all of the genome consisted of unique sequences then :

TABLE 4.1 KINETIC DATA

FRACTION	% OF GENOME	$COT_{\frac{1}{2}}$ OBS	$COT_{\frac{1}{2}}$ PURE[a]	CORRECTED $COT_{\frac{1}{2}}$ PURE		ANALYTICAL COMPLEXITY[d]	REPETITION FREQUENCY[e]
				%G+C[b]	MISMATCH[c]		
HIGHLY REPETITIVE + PALINDROMES	4.6	-	-	-	-	1.57×10^9	-
CLASS I REPETITIVE	43.3	6.8	2.94	2.646	1323	1.481×10^{10}	19,900
CLASS II REPETITIVE	20.4	2.45×10^3	500	450	N.D.	6.977×10^9	28[f]
SINGLE COPY	31.7	9.9×10^4	3.16×10^4	2.844×10^4	N.D.	1.084×10^{10}	0.68[f]

a) $Cot_{\frac{1}{2}}$ pure = $Cot_{\frac{1}{2}}$ obs x fraction of genome

b) Calculated from an empirical formula of Wetmur & Davidson (1968), assuming a %G+C of 46% for A. mexicanum and 50% for E.coli, assumes that the %G+C is similar for all classes.

c) Calculated from an empirical formula of Bonner et al (1973), assuming 9% mismatching (Fig. 4.4)

d) Analytical complexity = Genome size x fraction of genome

e) Repetition frequency = $\frac{\text{Analytical complexity}}{\text{Corrected } Cot_{\frac{1}{2}} \text{ pure}} \times \frac{Cot_{\frac{1}{2}} \text{ standard (single copy)}}{\text{Analytical complexity standard}}$

f) No correction for mismatching

g) N.D. Not determined.

$$\text{Cot}_{\frac{1}{2}} \text{ axolotl} = \frac{3.64 \times 10^{10} \times 8}{4.5 \times 10^{6}}$$

$$= 6.47 \times 10^{4}$$

The experimental observations clearly show that the majority of the axolotl genome can be defined as repetitive DNA at this criterion of reassociation. From Table 4.1 it is suggested that 68.3% of the axolotl genome reassociates faster than the expected single copy reassociation rate. This value however is probably an overestimate because even at short fragment lengths a proportion of the single copy fraction will be found adjacent to repetitive sequences, the proportion differing in degree depending on the average fragment length and the observed pattern of organisation of the genome. Hydroxylapatite chromatography, as discussed in the Appendix, does not discriminate between wholly or partially base paired duplexed so that single copy sequences contiguous to repetitive sequences will be scored as repetitive in this type of analysis.

The repetitive fraction of the genome has been reduced to three components, with the reservations already noted. A component of the genome was found to be double stranded at the earliest Cot points measured. The values obtained at early Cot do not markedly change over the range 10^{-5} to 10^{-1}. This rapidly reassociating material probably represents those regions of the genome in which adjacent inverted repeats can form a duplex of sufficient length to be stably bound to hydroxylapatite. The component is probably not satellite DNA for the following reasons :

a) In the previous section it was concluded that no major satellite fraction existed, although it must be stressed that the examination was not exhaustive;

b) divergence within satellite sequences tends to reduce the observed rate of reassociation and hence increase the $\text{Cot}_{\frac{1}{2}}$ for the reassociation (Southern, 1971; Hutton and Wetmur, 1971; Bonner

et al, 1973), so that satellite DNA should not reassociate before E Cot 10^{-5}, the point at which reassociation of this component appears complete.

None of the experiments shown excludes the possibility of a small component of the genome ($\leqslant$ 1%) existing as a satellite.

The largest component of the axolotl genome (43.3) has been termed Class I repetitive here. This class covers sequences reassociating within the Cot range E Cto 10^{-1} to 10^{+2}. The observed $Cot_{\frac{1}{2}}$ for the reassociation is 6.8 which by comparison to the kinetic standard, and after correction for the effects of % G+C content and mismatching (see Fig 4.4) on rate, yields an average repetition frequency of 19,900. As previously noted this figure represents the mean value for a range of families of various sizes and levels of divergence.

The smaller repetitive component (Class II, 20.4% of the genome) covers a Cot range of E Cot 10^{+2} to 10^{+4} and has an average repetition frequency of 28, although no estimate of the effect of divergence on rate has been made for this class. This class may include multiple gene families or repetitive families of low copy number. Sequences coding for ribosomal RNA and histones might be expected to be found in the Class I group (see for example Hilder et al, 1981).

The axolotl genome appears to consist mainly of repetitive families of varying degrees of repetition at this criterion of reassociation. The genome is qualitatively similar to the related ambystomatid salamander Ambystoma tigrinum (Straus, 1971). This animal has a similarly high C value and can be resolved into similar components. Estimates for these components and for those of other Urodeles from published data are given in Table 4.2. These estimates are uncorrected for % G+C or mismatching. In some cases the estimates have been deduced from figures rather than from stated $Cot_{\frac{1}{2}}$ values for components, these instances are noted in the figure.

TABLE 4.2

URODELE REASSOCIATION KINETIC DATA

SPECIES		C VALUE pg	HIGHLY REPETITIVE + FOLDBACK	%	REPETITIVE I $Cot\frac{1}{2}$ PURE	REPETITIVE I AVERAGE[a] REPETITION FREQUENCY	%	REPETITIVE II $Cot\frac{1}{2}$ PURE	REPETITIVE II AVERAGE[a] REPETITION FREQUENCY	%	UNIQUE $Cot\frac{1}{2}$ PURE	UNIQUE AVERAGE REPETITION FREQUENCY	REFERENCE
AMBYSTOMA													
MEXICANUM		37	4.6	43.3	2.94	8950	20.4	500	25	31.7	3.16×10^{4}	0.68	THIS THESIS
TIGRINUM	1	33	6.1	36.9	5.88	3480	27	379	54.1	30	N.D	N.D	STRAUS (1971)
	2b	27.5	5.7	72.3	N.D	-	-	-	-	22	N.D	-	MORESCALCHI (1980)
AMPHIUMA													
MEANS[b]		75	11.5	60.5	N.D	-	-	-	-	28	-	-	"
ANDRIAS													
JAPONICUS[b]		46.5	21	13	1.3×10^{-3}	-	41	0.53	-	10.2	-	-	"
DESMOGNATHUS													
FUSCUS[b]		15	34	27	8.6×10^{-2}	-	13	0.67	-	26			"

..... cont.

TABLE 4.2 (cont.)

SPECIES	C	VALUE pg	HIGHLY REPETITIVE + FOLDBACK	%	REPETITIVE I $Cot\frac{1}{2}$ PURE	REPETITIVE I AVERAGE[a] REPETITION FREQUENCY	%	REPETITIVE II $Cot\frac{1}{2}$ PURE	REPETITIVE II AVERAGE[a] REPETITION FREQUENCY	%	UNIQUE $Cot\frac{1}{2}$ PURE	UNIQUE AVERAGE REPETITION FREQUENCY	REFERENCE
NECTURUS													
MACULOSUS	1	52	13.6	54.1	9.4×10^{-2}	4.98×10^{4}	20.2	147	334	12.1	N.D	N.D	STRAUS (1971)
	2	52	20	47	0.423	4.15×10^{4}	10	15	1170	23	1.38×10^{4}	1	BALDARI & AMALDI (1976)
	3[b]	82.5	18	8	8×10^{-4}	-	49	0.49	-	25	-	-	MORESCALCHI (1980)
PLETHODON													
CINEREUS		20	15	30	0.42	9642	15	15	135	40	1.4×10^{4}	1	MIZUNO & MACGREGOR (1974)
TRITURUS													
CRISTATUS	1	23	10	43	0.6	1.2×10^{4}	-	-	-	47	7×10^{3}	1	BALDARI & AMALDI (1976)
CARNIFEX	2[d]	23	10	20			10			60	-	-	SOMMERVILLE & MALCOLM(1975)
	3[b]	23	9	20.5	0.06					69.5	-	-	MORESCALCHI (1980)
	4[c]	23	10	27	1	2×10^{4}	23	7.9	270	40	6.3×10^{3}	3.3	HILDER (unpublished)

..... cont.

TABLE 4.2 (cont.)

SPECIES	C VALUE pg	HIGHLY REPETITIVE + FOLDBACK	%	REPETITIVE I $Cot\frac{1}{2}$ PURE	REPETITIVE AVERAGE[a] REPETITION FREQUENCY	%	REPETITIVE II $Cot\frac{1}{2}$ PURE	REPETITIVE AVERAGE[a] REPETITION FREQUENCY	%	UNIQUE $Cot\frac{1}{2}$ PURE	UNIQUE AVERAGE REPETITION FREQUENCY	REFERENCE
TARICHA												
TAROSA[b]	28	10	20.5	0.01		36.5	1.25		33	-	-	MORESCALCHI (1980)

a. UNCORRECTED FOR DIVERGENCE AND %G+C CONTENT

b. VALUES DEDUCED FROM PUBLISHED DIAGRAMS

c. ANALYSIS BY S_1 NUCLEASE

d. ANALYSIS BY OPTICAL TECHNIQUES

N.D. NOT DETERMINED

The following points can be deduced from the combined data.

Most of the species listed show a fraction of the genome which may be termed single copy. Mizuno and MacGregor (1974) showed that for Plethodon cinereus cinereus the slowest reassociating component could be characterised by a $Cot_{\frac{1}{2}}$ similar to that predicted for single copy DNA under their conditions of reassociation. Similarly Baldari and Amaldi (1976) have identified fractions of the genome with the characteristics of single copy DNA in Triturus cristatus carnifex and Necturus maculosus.

Subject to two reservations it can be suggested that in general Urodeles with a large C value have a lower relative amount of single copy DNA than those with a lower C value. The two reservations are first that in some cases different authors reach different results for the same species so that comparisons between species need to be carefully assessed and secondly in most cases standard reassociation conditions have been employed so that the actual criterion of reassociation is different from species to species. However, as most Urodeles appear to have similar % G+C content where this has been measured (see Table 3.2), this effect is minor.

Even after noting these reservations the absolute amount of single copy DNA quite large, around 12 pg for the species noted in the Table 4.2. It has been estimated that less than 2% of the human genome can be accounted for as coding sequence based on estimates of genetic load (see for example, Salser and Isaacson, 1976), which suggests that the majority of human single copy DNA (50% of the genome, Schmid & Denninger, 1975) is non coding. (Lewin (1981) suggests that such arguments may however underestimate coding potential if a proportion of genes are not truly essential for survival, e.g. the w gene in D. melanogaster). Even so, it has been demonstrated in many cases that the sequence complexity of RNA in any given tissue is only a fraction of the single copy component of the genome (Davidson, 1976). It seems unlikely that Urodeles would require

additional coding capacity, so that most of the single copy DNA of Urodeles may be informationally redundant. Rosbash et al (1974) have shown that the sequence complexity expressed in ovary poly A^+ RNA is the same in Xenopus and Triturus while Mizuno and MacGregor (1974) showed that only 40% of the single copy fraction is conserved between species of the genome Plethodon. The large amounts of single copy DNA in Urodeles would appear to be noncoding. It is most likely that the majority represents repetitive sequences which cannot reassociate at the criterion of reassociation used. (See also Graham and Schanke, 1980; Murray et al, 1978; Bouchard and Swift, 1977; Moyzis et al, 1981 b).

As the C value increases the proportion of the genome that is repetitive increases. This can be seen from the data in Table 4.2. This correlation has been observed by several groups working on both animal and plant genomes. (Baldari and Amaldi, 1976; Flavell et al, 1974; Mizuno and MacGregor, 1974). The increase can be achieved by an increase in the repetition frequency of existing families and hence little increase in the sequence complexity within this fraction, or by the formation of new repetitive families. Baldari and Amaldi (1976) have concluded that for Triturus cristatus carnifex and Necturus maculosus the increase in the repetitive fraction may be due to a general increase in repetition frequency, although they do not discount the formation of new repeat families. A comparison of the uncorrected data for the axolotl with that of the related salamander A. tigrinum might also suggest that the difference between the two species is due to expansion of existing families. However, it is also possible to postulate the formation of new repeat families. Mizuno and MacGregor (1974) have shown that species of the genus Plethodon which vary in C value and degree of repetition have few repeats in common, so that even closely related species have generated new repeat families. An analogous situation might exist between the Ambystomatids.

As suggested in the introduction there are many instances of both

novel family formation and repeat family remultiplication and models for such events have been put forward (Klein et al, 1978; Flavell, 1981). In large genomes these processes can be hypothesised to be occuring relatively frequently so it may be possible to isolate examples of the various types of family organisation . The general organisation of the axolotl genome as revealed by the reassociation data shows similarities to other high C value Urodeles. However such data do not reveal the relative organisation of each sequence class within the genome. The interspersion analysis of Davidson et al (1973) has been used to determine the relative sequence organisation in the axolotl genome.

Briefly, radioactively labelled DNA of various single strand sizes are mixed with a large excess of short unlabelled DNA fragments. These are denatured and allowed to reassociate until the sequence class under study in the driver DNA has completely reassociated. Aliquots are removed and completely single stranded DNA is separated from duplex containing material by hydroxylapatite chromatography. In this way the tracer DNA containing the sequence class of interest becomes bound to hydroxylapatite. As the tracer length increases the sequences adjacent to the class under study are also bound. When the fragment length is increased such that it is longer than the sequence length intervening between two members of the class under study, then no extra label will be bound. This is summarised in Fig. 4.2.

One requirement for this assay is that the sequence class being studied should be fully renatured before the next class begins to reassociate in order to avoid the effects of tracer length on reassociation rate (See Appendix II; Hinnebusch et al, 1978; Moyzis et al, 1981 a + b). From Fig. 4.1 it can be seen that the majority (95%) of Class I repeats have completed reassociation by E Cot 100. The Class II repeats do not begin to reassociate until an E Cot of around 500. Therefore it is possible to distinguish between the reassociation of the Class I repetitive sequences,

LEGEND TO FIG. 4.2

SCHEMATIC REPRESENTATION OF SEQUENCE INTERSPERSION

In this type of experiment radiolabelled DNA (tracer DNA) of varying single strand size is renatured in the presence of a vast excess of short unlabelled DNA (driver DNA), to a Cot value at which most or all of the repeat class of interest has renatured, but at which little or none of the slower reassociating class(es) have reassociated. Duplexes are bound to hydroxylapatite and the fraction bound (corrected for binding of palindromic sequences (Davidson et al, 1973)) is plotted vs tracer single strand size.

a) No interspersion : If all of the repetitive sequences of interest are clustered and not interspersed with slower reassociating sequences increasing the length of the tracer DNA will not increase the fraction bound after reassociation to a Cot point at which the repeats are in duplex form.

b) Interspersion - Short and long period : The interspersion period is the distance between two sequences of the repeat class under study along a contiguous stretch of sequence. Short period interspersion is characterised by the presence of more than one repeat sequence per fragment at moderate fragment lengths. Long period interspersion is characterised by the presence of one repeat sequence only on relatively long fragments.

The presence of both patterns gives rise to a curve qualitatively similar to (b). See main text for more details.

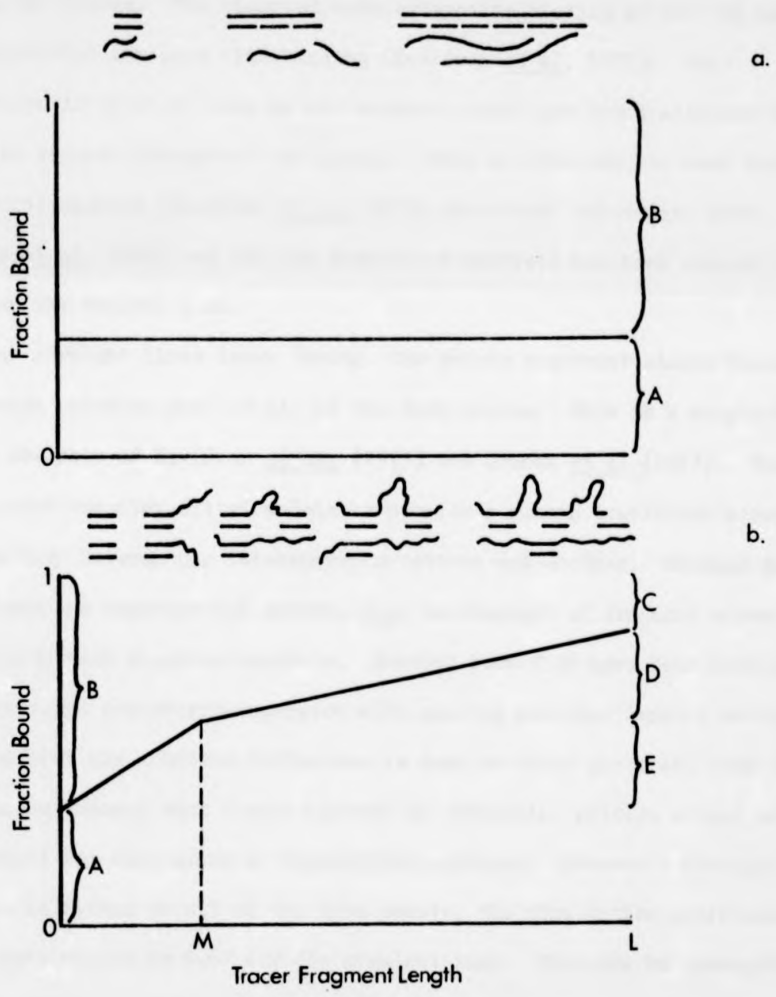
A. Fraction of the genome in the repetition class of interest.

B. Fraction of the genome in a more slowly reassociating class(es).

C. Fraction of the genome not adjacent to a member of the repetitive class of interest at fragment length, L.

D. Fraction of the genome adjacent to one repetitive sequence at fragment length, L.

E. Fraction of the genome, of average length M, adjacent to one repetitive sequence.


a.
1
Fraction Bound
0
B
A
b.
1
Fraction Bound
0
B
A
C
D
E
M
L
Tracer Fragment Length

the largest repetitive group in the genome, and the more slowly reassociating Class II repetitive sequences and single copy sequences. Therefore it has been possible to analyse this class in some detail. Figure 4.3 shows the data on interspersion of Class I repetitive sequences within the genome. The observed hydroxylapatite binding at Cot 100 has been corrected for zero time binding (Davidson et al, 1973). This correction is valid as long as all sequence types are found adjacent to inverted repeats throughout the genome. Such a situation has been found in several species (Davidson et al, 1973; Schmid and Deininger, 1975; Perlman et al, 1976) and for the purpose of analysis has been assumed to occur in the axolotl also.

The straight lines drawn through the points represent simple linear regression covering part or all of the data points. This is a simplification of the analysis of Davidson et al, (1973) and Graham et al (1974). These groups used computer fitted models to provide a smooth transition across the boundary between one interspersion pattern and another. However for most cases the experimental errors, e.g. measurement of fragment sizes, do not justify such rigorous analyses. Several best fits have been plotted assuming short period interspersion with spacing sequence lengths averaging 0.6, 0.8 or 1 kbp. Little difference is seen in these plots and from this it could be assumed that a short period interspersion pattern exists in the axolotl for this class of repetitive sequences. However a straight line can be fitted to all of the data points, the correlation coefficient of the data points is 0.67 for the straight line. This can be interpreted to suggest that a long period pattern of interspersion exists for this class of repeats, the distance between repeats being at least 4 kb. This interpretation is greatly influenced by the observed scatter of the points. As will be discussed later the short period interspersion is favoured.

Recently Moyzis et al (1981 a + b) have proposed an alternative analysis of such data. They suggest that it is possible to re-evaluate much of the

existing sequence organisation data, and that the short period interspersion pattern observed by many authors may be only one of several equally plausible interpretations of the data. Indeed Graham et al (1974) stress that their original data fitting was done in such a way as to simplify the calculations (p.133). The original interpretations were influenced by contemporary ideas about gene structure based on mRNA length and on the possibility of coordinate expression of multigene "batteries" controlled by the actions of trans acting regulator molecules on adjacent (repetitive) sequences (see for example Britten and Davidson, 1969). However the discovery of intervening sequences within gene coding regions, and detailed analysis of recombinant DNA clones of genomic DNA containing gene sequences suggests that a different interpretation may be required. This does not mean that repetitive sequences are not found around coding regions, there is a large body of experimental data supporting this arrangement (e.g. Davidson, 1976; Davidson and Posakony, 1982; Bishop and Freeman, 1974; Hacket and Lis, 1981; Crampton et al, 1981; Ryffel et al, 1981; Zuker and Lodish, 1981; Grosschedl and Birnstiel, 1980; Kimmel and Firtel, 1980). However the gene coding fraction of the genome is quite small (see earlier) and the organisation of these sequences and surrounding regions need bear no relation to the organisation of the majority of the genome. (See also Davidson et al, 1977).

Bearing these considerations in mind several conclusions can be drawn from the data in Fig. 4.3a. By comparison to the schematic diagram in Fig. 4.2 it can be seen that between 42.5 and 48.2% of the genome belongs to the Class I repetitive fraction, depending on the fit to the data. The estimate from Table 4.1 of 43.3% is in good agreement with these figures and tends to suggest that a proportion of the Class is interspersed with more slowly reassociating sequences. However it is clear that up to fragment lengths of 4 kbp the majority of sequences in this Class are interspersed with other sequences of the same class. This can be deduced

LEGEND TO FIG. 4.3

HYDROXYLAPATITE BINDING WITH INCREASING TRACER FRAGMENT SIZE

^{3}H labelled DNA of various single strand sizes was prepared as Methods.

a) Cot 100 Assay. ^{3}H DNA was used as a tracer in reassociations driven by a vast excess (greater than 1000:1) of short genomic driver DNA. DNA was allowed to reassociate in 0.41M NaPB pH6.8 at 69°C to driver E Cot of 100, calculated according to Britten et al (1973). After reassociation the solutions were diluted to 0.12M NaPB using ice cold distilled water, before hydroxylapatite chromatography.

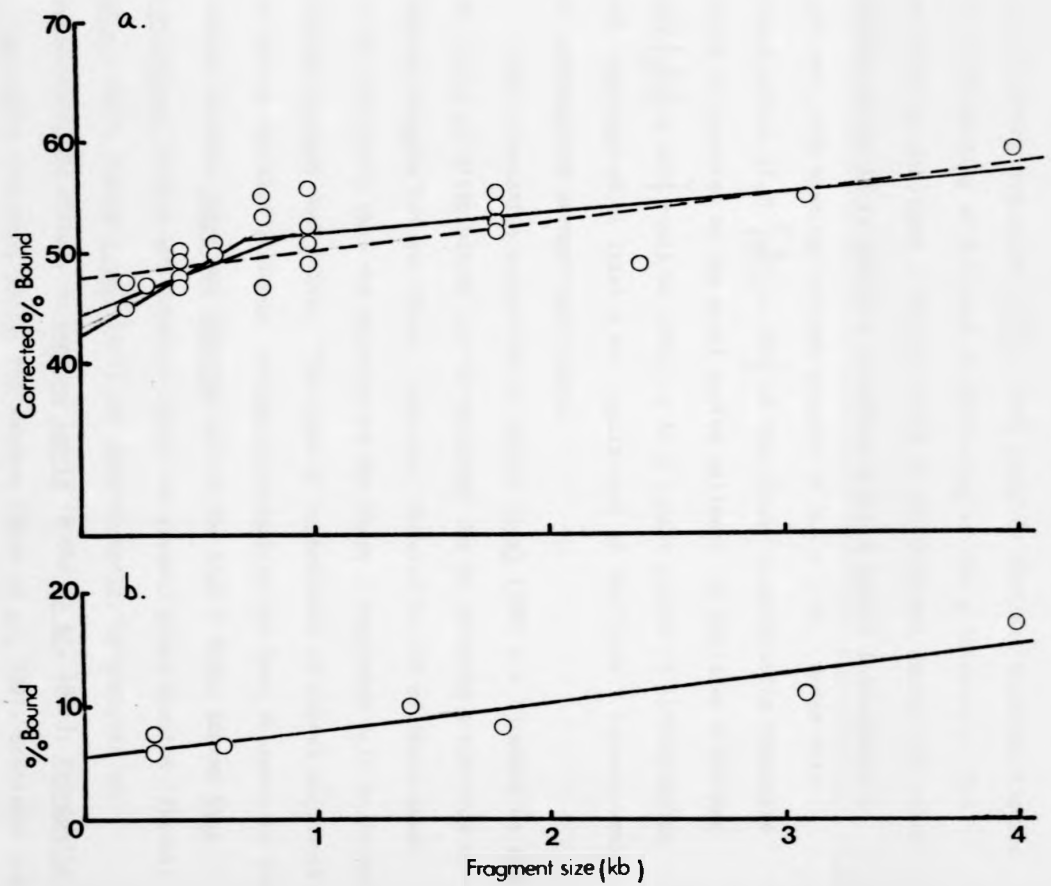
The hydroxylapatite bound fraction at each fragment size was corrected for zerotime binding (b) by

$$\text{Corrected binding} = \frac{\text{Fraction bound at length,l-zerotime binding at length,l}}{\text{1-zerotime binding at length,l}}$$

according to Davidson et al (1973). Best fits to the data are described in the text.

b) Zerotime binding. 1-5 ng of ^{3}H labelled DNA was mixed with 50 ug sheared E.coli DNA in 1 ml 0.12M NaPB pH6.8. The solutions were boiled 5' at 100°C, quenched on ice and immediately assayed on hydroxylapatite at 60°C. The highest attainable Cot for the H^{3} DNA was calculated to be 5×10^{-6}.

The line through the data points is a simple best fit.


a.
70
60
50
40
Corrected % Bound
b.
20
10
0
% Bound
1
2
3
4
Fragment size (kb)

from the slow rate of increase of binding to hydroxylapatite with increasing fragment length.

The data suggest at least two possible patterns of arrangement for the genome involving these sequences. The straight line fit to the data suggests that a fraction of the genome is arranged such that sequences of at least 4 kb are adjacent to Class I repeats. 17.9% of the slower reassociating sequences $\left\{\frac{9.3}{51.8} \times 100\right\}$ would be involved based on figures of 57.5% binding at 4 kb and 48.2% binding at the y intercept. The majority of the Class I repeats would be interspersed among each other. Alternatively it is possible to deduce a short period interspersion pattern, with spacings between repeats of 0.6 - 1 kb. Under this organisation 13.9% $\left\{\frac{8}{57.5} \times 100\right\}$ of the slower reassociating sequences would be involved in the short period pattern. In addition a further 95% $\left\{\frac{5.5}{57.5} \times 100\right\}$ would be involved in a longer period of interspersion with spacings of at least 4 kb. Again most of the Class I repeats would be interspersed amongst each other.

The alternative suggestion of Moyzis et al (1981 a + b) would be that this kind of distribution can be accounted for by assuming a spectrum of spacing lengths between Class I repeats. However in all of these cases it is inevitable that the majority of the Class I sequences will be interspersed amongst each other. This type of arrangement of repeat sequences is not an isolated instance. A similar situation has been observed in the related Urodele Ambystoma tigrinum and in the high C value Anuran Rana berlandieri (Graham and Schanke, 1980), in several plant species (Flavell et al, 1977; Murray et al, 1978) and most recently, by analysis of recombinant DNA clones, in Xenopus laevis (Spohr et al, 1981), Drosophila melanogaster (Wensink, 1977), the chicken (Eden et al, 1981; Sob ieski and Eden, 1981; Musti et al, 1981) and the sea urchin Strongylocentrotus purpuratus (Anderson et al, 1981). The arrangement can also be inferred by the presence in most animal and plant genomes of long S_1 nuclease

resistant repetitive DNA isolated after reassociation of moderately sheared DNA to intermediate Cots (see for example, Galau et al, 1976; Pearson et al, 1978; Moyzis et al, 1981 b). The possible development of such sequences has already been described in the Introduction.

Fig. 4.3b shows the binding of DNA at Cot 10^{-6}. At such early Cot points only adjacent inverted repeats will have renatured. The increase in binding with fragment length is apparently linear over the fragment range analysed. Extrapolating to the y axis indicates that 5.5% of the genome is arranged as inverted repeats. At fragment lengths of 4 kb only 15% of the genome bound to hydroxylapatite at Cot 10^{-6}. The rate of increase of binding with fragment length, together with the estimate of the fraction of the genome which occurs as inverted repeats, suggests that foldback sequences are clustered in the genome. These clusters may themselves be spread randomly through the genome, as suggested for Triturus (Wilson and Thomas, 1974).

It has been shown by many groups that the repetitive sequence fraction of the genome displays varying degrees of homology in its reassociated products. This can be demonstrated by controlled thermal denaturation of the reassociated fraction. In most instances a broad spread of thermal stability can be observed in the order of 8 - 10°C. It has been shown, by Bonner et al (1973) that a 1°C reduction in Tm corresponds to 1% mismatch between the reassociated strands. Thus most repetitive sequences show an average divergence of around 8 - 10%, although exceptions do occur, (Wensink, 1977; Crain et al, 1976). The melting behaviour of the Class I repeats was examined in similar fashion and the results are plotted in Fig. 4.4.

Sheared native axolotl DNA showed a T_E (point of 50% elution from HAP) of 86°C, whereas the same DNA denatured and allowed to reassociate to E Cot 100 showed a T_E of 76°C, a reduction of 10°C. By the criterion of Bonner et al (1973) this implies an average sequence divergence of 10%

LEGEND TO FIG. 4.4

MELTING CURVES OF SHEARED DNA

Genomic DNA was ^{32}P labelled by nick translation. 1 ng (5×10^3 Cerenkov cpm) was mixed with 50 µg of 300 b cold DNA in 0.12M NaPB for each assay.

Samples were used directly (native) or denatured and renatured to Cot 100 (Cot 100). 50 ug aliquots were loaded onto 1 ml HAP in 0.12M NaPB at 60°C. Unbound material was washed from the column with 3 x 5 ml washes. The temperature of the column was raised in 5°C increments. The column was held for 5' at each new temperature, then washed with 2 x 5 mls of 0.12M NaPB preheated to the elevated temperature. At 98°C (the highest practicable temperature) the column was washed with 2 x 5 mls 0.12M NaPB then 2 x 5 mls 0.4M NaPB to release any remaining bound counts. These elution conditions have been shown by Martinson (1973 a + b) to cause immediate elution upon denaturation, and also to prevent desorption of still duplexed molecules.

Irreversibly bound counts were measured by dissolving the HAP in 25% TCA and counting. The amount irreversibly bound was usually less than 1%.

The data was plotted as (a) Percent eluted vs temperature or
(b) Fraction eluted at each increment vs temperature.

a) Percent eluted vs temperature.

Cot 100 bound, Tm = 76 (Δ)

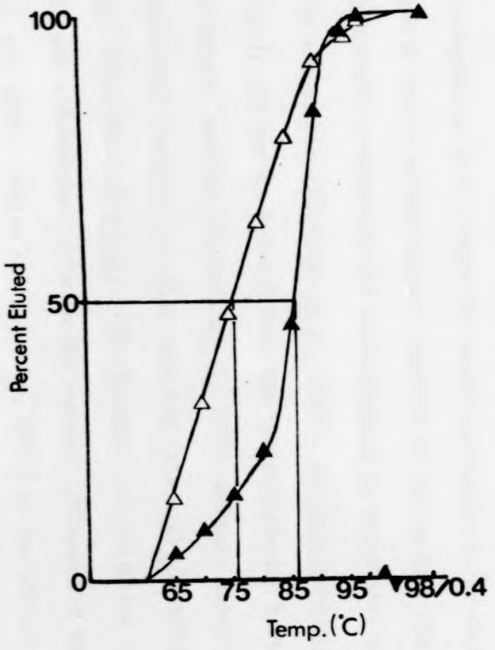
Native bound, Tm = 86 (▲)

b) Fraction eluted/increment.

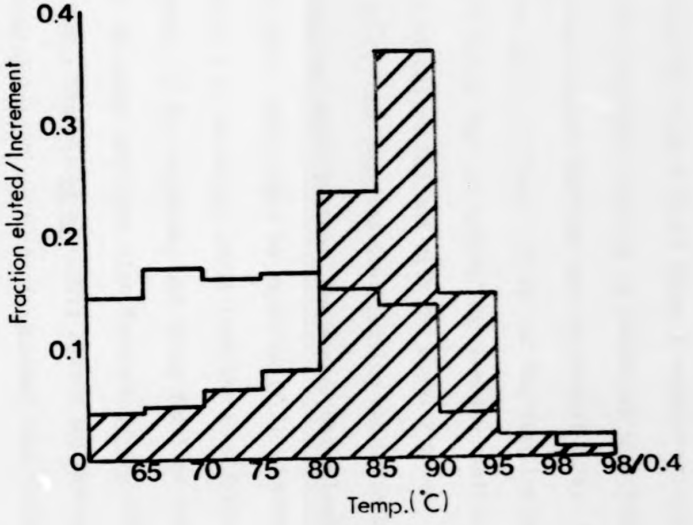
Open histogram Cot 100

Hatched histogram Native

a.


100
50
0
Percent Eluted
65
75
85
95
98/0.4
Temp.(°C)

b.


0.4
0.3
0.2
0.1
0
Fraction eluted / Increment
65
70
75
80
85
90
95
98
98/0.4
Temp.(°C)

for the whole class of repeats. It can be seen from the plot of fraction eluted at each temperature (hatched Histogram) that few repeats show high stability (or alternatively few high % (G+C) Class I repeats exist), however below 90°C a fairly constant fraction is eluted at each temperature. This suggests that different repeat families have different T_E's. The reason for this suggestion is as follows. If all of the families had the same T_E and this was 10°C below that of native DNA then the elution pattern would be similar to that of native DNA but with the modal elution temperature shifted by 10°C. The observed distribution can be accounted for by a series of overlapping distributions, the sum of which leads to the constant elution profile. This might be expected if it is assumed that repeat families are produced by saltatory amplifications at different points in the evolutionary history of the organism, and these families then accumulate mutations at the same rate (see also Thompson et al, 1980; Thompson and Priesler, 1981; Flavell et al, 1977; Mizuno and MacGregor, 1974). Thus "inter familial heterogeneity" can be distinguished from "intra familial heterogeneity" (Bouchard and Swift, 1977).

Although it is possible to distinguish between these two extremes of organisation it is not possible to determine whether all of the repeat families are homogeneous with respect to their particular Tm or whether microheterogeneity exists within repeat families in the axolotl Class I repeats. Such microheterogeneity has been observed in repetitive sequence families in the sea urchin (Scheller et al, 1981), Drosophila melanogaster (Lis et al, 1981), and can be deduced from the reamplification history of certain plant repeat families (Flavell et al, 1977) and many satellite species. Homogeneous families are also observed (Klein et al, 1978; Rinehart et al, 1981; Wensink et al, 1978; Thompson and Priesler, 1981) which may indicate a functional role in the genome, (Mizuno and MacGregor, 1974; Klein et al, 1978; Bendich and Anderson, 1977) or the effect of sequence conversion (Dover, 1981).

for the whole class of repeats. It can be seen from the plot of fraction eluted at each temperature (hatched Histogram) that few repeats show high stability (or alternatively few high % (G+C) Class I repeats exist), however below 90°C a fairly constant fraction is eluted at each temperature. This suggests that different repeat families have different T_E's. The reason for this suggestion is as follows. If all of the families had the same T_E and this was 10°C below that of native DNA then the elution pattern would be similar to that of native DNA but with the modal elution temperature shifted by 10°C. The observed distribution can be accounted for by a series of overlapping distributions, the sum of which leads to the constant elution profile. This might be expected if it is assumed that repeat families are produced by saltatory amplifications at different points in the evolutionary history of the organism, and these families then accumulate mutations at the same rate (see also Thompson et al, 1980; Thompson and Priesler, 1981; Flavell et al, 1977; Mizuno and MacGregor, 1974). Thus "inter familial heterogeneity" can be distinguished from "intra familial heterogeneity" (Bouchard and Swift, 1977).

Although it is possible to distinguish between these two extremes of organisation it is not possible to determine whether all of the repeat families are homogeneous with respect to their particular Tm or whether microheterogeneity exists within repeat families in the axolotl Class I repeats. Such microheterogeneity has been observed in repetitive sequence families in the sea urchin (Scheller et al, 1981), Drosophila melanogaster (Lis et al, 1981), and can be deduced from the reamplification history of certain plant repeat families (Flavell et al, 1977) and many satellite species. Homogeneous families are also observed (Klein et al, 1978; Rinehart et al, 1981; Wensink et al, 1978; Thompson and Priesler, 1981) which may indicate a functional role in the genome, (Mizuno and MacGregor, 1974; Klein et al, 1978; Bendich and Anderson, 1977) or the effect of sequence conversion (Dover, 1981).

Apart from the deductions noted above for the axolotl, data from cloned repetitive sequences indicate that inter familial heterogeneity does occur in the axolotl (See Chapter 5). The consequences of inter familial heterogeneity in repeat classes has been discussed in detail by Bendich and Anderson (1977) amongst others with respect to the relationship between the criterion of reassociation and the observed sizes of the various repetition classes within the genome (e.g. Bouchard and Swift, 1977; Murray et al, 1978; Thompson et al, 1980).

Earlier it was deduced from the interspersion data that the majority of the Class I repeats are not interspersed amongst more slowly reassociating sequences under the conditions used. The interspersion of Class I repeats amongst each other can occur in several ways. Two extremes of organisation are first , that long tandem repeat arrays exist in which each member of a particular repeat family is adjacent to similar sequences or second, that all family members are dispersed amongst other repeat elements of different origin, length, degree of divergence, % G+C, methylation. Between these two extremes it is possible to define a multiplicity of possible arrangements. In earlier sections results suggest that repetitive sequences in the axolotl may be both interspersed (Fig 3.6) while some degree of tandem repetition or clustering of similar sequence types may occur (Figs. 3.4 and 3.7). In order to try and distinguish to some extent between these possibilities the following experiments were performed. Native in vivo ^{3}H labelled DNA (weight average double strand length $\geqslant$ 50k base pairs) was denatured and reassociated to various observed Cot values. The single stranded regions were removed by digestion with the single strand specific S_1 nuclease under conditions in which $\geqslant$95% of single stranded DNA is digested, but double stranded DNA remains essentially undigested (see Methods). The duplexes were fractionated according to size and the size distribution was plotted as Fig. 4.5. In this figure two Cot points, 0.01 and 100 are shown. Several points are immediately noticeable.

LEGEND TO FIG. 4.5

SIZE DISTRIBUTION OF DUPLEX DNA

Native, in vivo ^{3}H labelled DNA (1.5×10^{3} cpm/ ug) was denatured and reassociated to various observed Cot values. The DNA was quick frozen on dry ice/ethanol to stop the reassociation. Aliquots were taken and treated with S_1 nuclease to remove single strand tails. Sufficient S_1 nuclease was added to digest 95% of single stranded DNA (see Methods). The S_1 nuclease resistant DNA was recovered and a portion taken to estimate % TCA precipitable counts. The main volume of each reassociation was mixed with 2 ug of pBR322 which had been digested with HinfI (Sutcliffe, 1978). The DNA's were fractionated by size on 2% agarose tube gels. The gels were sliced into 2 mm slices. The slices were dissolved in formamide at 60°C cooled and counted in an acidified scintillation cocktail (1 ml glacial acetic acid per litre of Triton-Toluene-PPO-POPOP).

a) Cot 10^{-2}

b) Cot 100

Calculations : From Fig. 4.3 8% of the genome appears as short sequences of around 1 kb between two Class I repeats, while at least 5.5% of the genome is composed of sequences at least 4 kb.long adjacent to one Class I repeat. The genome size is 3.4×10^{10} base pairs. To account for the interspersion 2.8% of the Class I repeats would be required. The calculation is as follows :

i. Short interspersion

$$0.08 \times \frac{3.4 \times 10^{10}}{1000} = \text{no. of interspersed sequences in short period pattern}$$

$$= 2.72 \times 10^{6}$$

ii. Long interspersion

$$0.055 \times \frac{3.4 \times 10^{10}}{4000} = \text{no. of interspersed sequences in long period pattern}$$

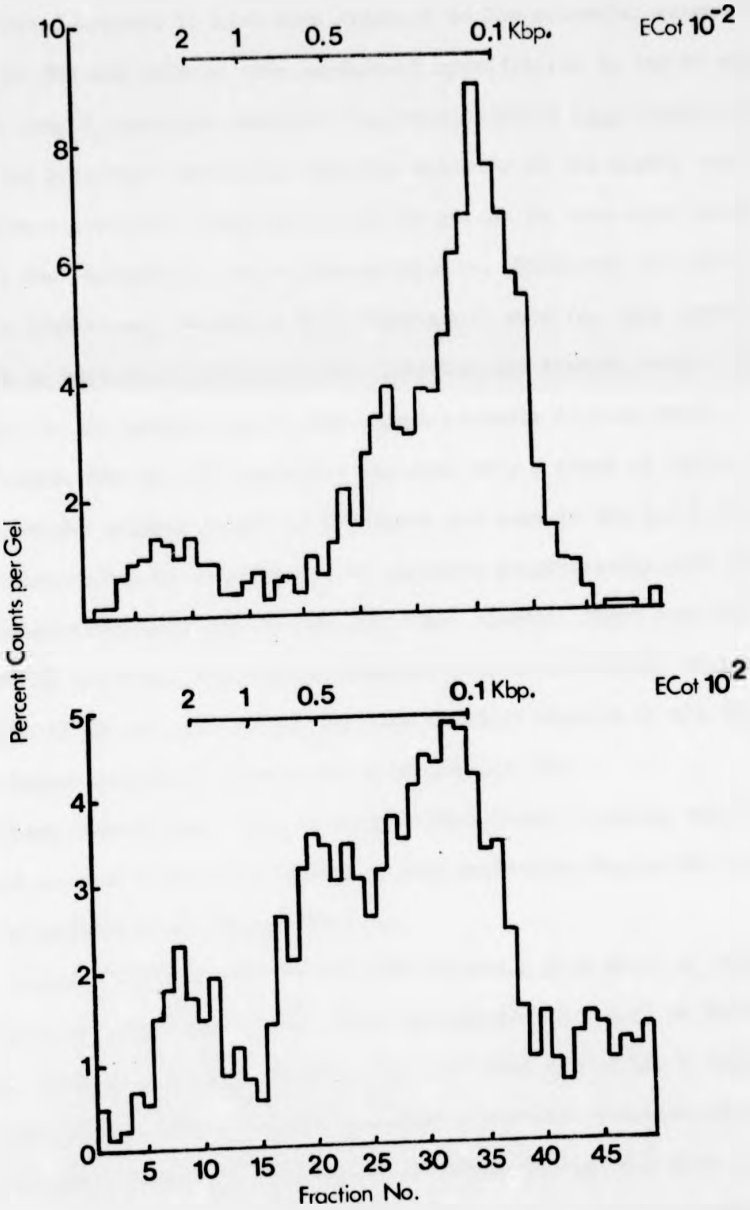
$$= 4.675 \times 10^{5}$$

Therefore $2.72 \times 10^{6} + 4.675 \times 10^{5}$ adjacent repetitive sequences are required to produce the observed interspersion.

Assuming a repeat length of 300 bp (Fig. 4.5) then the fraction of the genome required would be

$$2.72 \times 10^{6} + 4.675 \times 10^{5} \times \frac{300}{3.4 \times 10^{10}}$$

$$= \underline{0.028}$$


2 1 0.5 0.1 Kbp.
ECot 10^{-2}
10
8
6
4
2
Percent Counts per Gel
2 1 0.5 0.1 Kbp.
ECot 10^{2}
5
4
3
2
1
0
5 10 15 20 25 30 35 40 45
Fraction No.

First, in both of the distributions shown very little of the nuclease resistant material is much larger than 2 k bases in length, while most of the material appears to have been digested to low molecular weight duplexes. Unsheared DNA was used in this experiment specifically to try to establish whether long S_1 nuclease resistant duplexes do exist (e.g. satellites). Under the digestion conditions used the majority of the highly repeated S_1 nuclease resistant fraction (8% of the genome in this experiment) do not exhibit the characteristics of tandem repeats. Similarly the Class I repeats (48% of the genome in this experiment) show few long repeats. As the DNA is uniformly labelled these distributions measure weight average lengths, so the proportion of long repeat elements is very small.

Second, the Cot 100 resistant duplexes show a class of repeat lengths with a weight average length of 500 bases not seen in the Cot 0.01 fraction. This fraction can be demonstrated to increase progressively with increasing Cot, between Cot 0.01 and Cot 100 (data not shown). These duplexes presumably represent the average fragment lengths of Cot 100 repeats, although it should be stressed that the absolute amounts of all the fragment size classes increases between Cot 0.01 and Cot 100.

These observations tend to suggest that Class I repeats are interspersed in such a way that members of any particular family are located between members of different families.

Around 80% of the axolotl Cot 100 sequences form short S_1 nuclease resistant repeats (Fig. 4.5b). This represents 0.8 x 0.43 or 34.4% of the axolotl genome, and might therefore suggest that 34% of the axolotl genome consists of Cot 100 repeats surrounded by a variable fraction of slower reassociating sequences. However it was shown in Fig. 4.3 that 8% of the genome was slower reassociating sequences of around 1 kbp in length adjacent to Class I repeats, while at least 5.5% of the genome consisted of slower reassociating sequences at least 4 kbp long adjacent to a Class I repeat. Therefore, taking 300 bp as a reasonable estimate for the average length of

the short S_1 nuclease resistant repeats it can be estimated that a maximum of 2.8% of the genome (6.4%, (0.028 x 0.43) of the Class I repeats) need to be associated with slower reassociating sequences to account for the observations of Fig. 4.3. Of course this figure varies with the various assumptions made, being reduced as estimates of spacing lengths increase (and vice versa) and being increased as the estimate of the repeat length increases (and vice versa). Therefore an observable discrepancy exists between the observed 34.4% of the genome present as short repeats, and the value of 2.8% required, as a maximum, to account for the observed interspersion pattern. This discrepancy can be easily explained as being the result of reassociation of Class I repeats interspersed among themselves leading to the formation of accessible regions for S_1 nuclease action. These regions could occur for several reasons. Steric hindrance within networks of interspersed repeated sequences may prevent the reassociation of a proportion of repeats *e.g.* those Cot 100 repeats with the lower copy numbers in the repeat range. Reassociation between members of divergent repeats may present S_1 accessible regions within the reassociated duplexes, Posakony *et al* (1981) have shown by DNA sequencing of members of sea urchin repeat families that divergence may be due to small deletions, insertions and rearrangements (which may be S_1 nuclease accessible in stringent digestion conditions) as well as to simple base substitutions.

Recent work using recombinant DNA clones containing long repeat elements has confirmed that in many cases the long S_1 resistant repeats may consist of arrays of shorter repeats, these shorter repeats originating from different repeat families. This has been clearly shown in *Drosophila melanogaster* (Wensink, 1977), *Xenopus laevis* (Spohr *et al*, 1981), the sea urchin *Strongylocentrotus purpuratus* (Anderson *et al*, 1981; Scheller *et al*, 1981), the chicken *Gallus gallus* (Musti *et al*, 1981; Eden *et al*, 1981; Sobieski and Eden, 1981). Detailed analysis of the long S_1 resistant repeats in the Syrian hamster (Moyzis *et al*, 1981 a + b) and the amphibians

Ambystoma tigrinum and Rana berlandieri (Graham and Schanke, 1980) shows that the majority of these sequences are composed of shorter repeats. Moyzis et al, (1981 b), show that for the Syrian hamster most repetitive DNA sequences are present in both long and short S_1 nuclease resistant repetitive DNA duplexes, at approximately the same concentrations, Graham and Schanke (1980) have shown that the long S_1 resistant repeats in A. tigrinum and R. berlandieri are criterion sensitive suggesting that shorter repeat families with different degrees of divergence are interspersed amongst each other. This situation probably occurs in the axolotl and it may be proposed that a proportion, in some cases a substantial portion, of the short S_1 nuclease resistant repeats observed in many species may be of the same source.

An alternative method of analysing the organisation of a particular genome is to shear the DNA to moderate fragment lengths of several thousand bases (single stranded) and allow the DNA to reassociate. The fraction bound at each Cot point is determined by the reassociation of the most highly repetitive component present on each fragment. Hence by comparison with the reassociation of short DNA fragments relative interspersion patterns can be deduced. To reduce errors in analysis caused by networks of DNA being retained on hydroxylapatite the duplex DNA is eluted at 95°C (Kiper and Herzfeld, 1978).

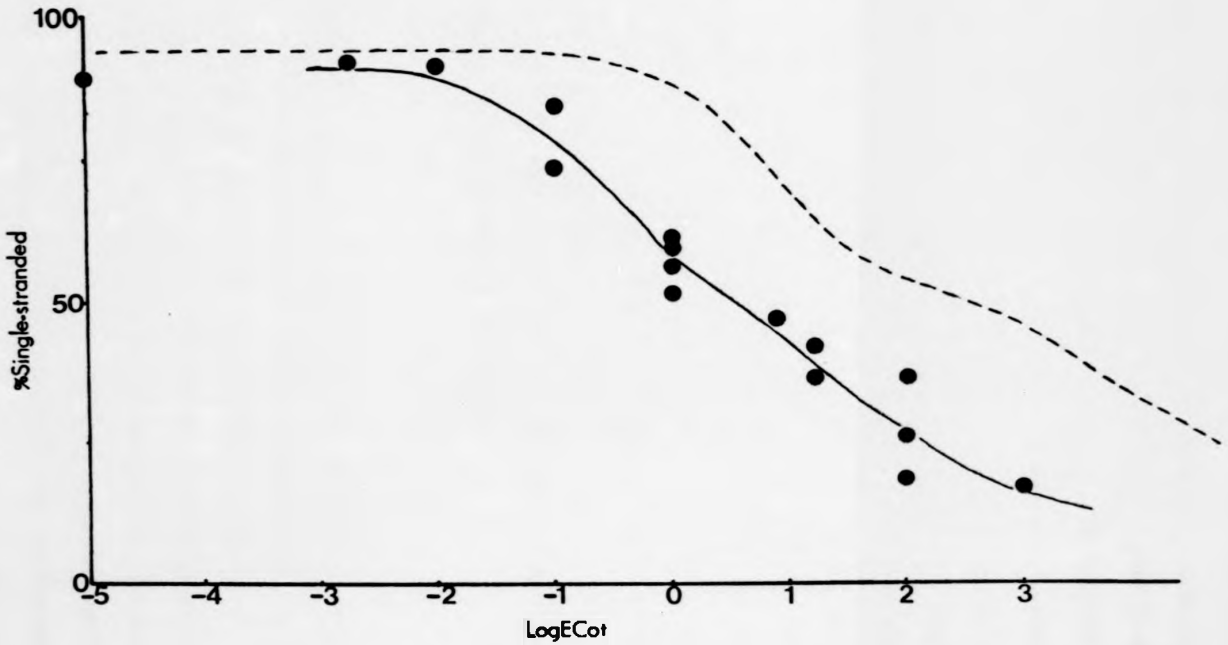
When axolotl DNA, sheared to a weight average single strand length of 2.5 kb was allowed to reassociate and was fractionated as described, the Cot curve shown in Fig. 4.6 was generated. Several features are apparent. The proportion of the genome assigned to each sequence class is altered compared to the curve shown in Fig. 4.1 (the fitted curve from Fig. 4.1 is shown in Fig. 4.6 as the broken line). Each of the repetitive fractions has increased in proportion in the genome. The highly repetitive fraction has increased from 5 to 9% of the genome. This value is similar to that obtained from Fig. 4.3b for 2.5 kb DNA. The major reassociating class, which most probably

LEGEND TO FIG. 4.6

2.5 kb REASSOCIATION KINETICS

Axolotl genomic DNA was sheared by sonication to a weight average single strand length of 2.5 kb. The DNA was denatured and renatured to various E Cot values. 50 μg aliquots were removed and fractionated by hydroxylapatite chromatography. Duplex DNA was removed by raising the column temperature to 95°C and eluting in 0.4M NaPB pH6.8.

The fit to the data was produced by computer analysis similar to that of Fig. 4.1.


100
50
0
%Single-stranded
-5
-4
-3
-2
-1
0
1
2
3
LogECot

corresponds to the Class I repeats seen in Fig. 4.1 comprise 50.4% of the genome in Fig. 4.6. The small increase in the apparent "size" of this class indicates that only a minor fraction of these repeats are interspersed amongst slower reassociating sequences. A second repetitive fraction, corresponding to the Class II repeats occupies 30.4% of the genome. This represents an apparent increase of 10% compared to the reassociation of 500 base DNA fragments. The slowest reassociating class (single copy) occupies a maximum of 15% of the genome at 2.5 kb. This reduction from around 32% at 500 b. indicates that while only approximately half of the sequences appear to be interspersed with more repetitive elements on 2.5 kb fragments, most of these are adjacent to the most slowly reassociating repetitive elements. It must be emphasised that while these values are based on fits to the data, the small number of points and the observed scatter tend to reduce the accuracy of the fit so that the assigned values are not necessarily the only ones which could be applied although the major conclusions would not be altered.

It can be concluded that in general the Class I repeats tend to be interspersed amongst other Class I repeats.

SUMMARY

Analysis of the axolotl genome by reassociation techniques has revealed the organisation of the various repetitive classes within the genome. The genome can be divided into several classes :

1) A highly repetitive/foldback fraction, comprising 46% of the genome. Sequences in this class appear to be clustered in the genome.

2) A repetitive fraction (Class I repetitive) which comprises 43.3% of the genome. Based on classical interpretation of interspersion experiments a proportion of this class of repeats appears to be organised in both long and short period interspersion patterns with sequences of lower average repetition frequency. The majority of Class I repeats appear to be

interspersed amongst repeats of the same Class.

The average repetition frequency is 19,900 corrected for an observed average divergence of 10%. The divergence pattern indicates extensive interfamilial heterogeneity in degree of divergence.

3) A second repetitive class (Class II) comprises 20.4% of the genome and has an uncorrected repetition frequency of 28.

4) The final fraction, 31.7% of the genome is probably single copy sequences. At fragment lengths of 2.5 kb half of these sequences appear to be adjacent to more rapidly reassociating sequences.

CHAPTER 5

RESULTS AND DISCUSSION III

MOLECULAR CLONING AND ANALYSIS OF AXOLOTL REPETITIVE DNA

The purpose of this section is to describe the formation and analysis of recombinant clones between genomic DNA and a plasmid vector. The clones will be analysed in terms of restriction enzyme mapping, reassociation kinetics and genomic organisation. The relative divergence of each clone will be determined and the transcription of members of these repeat families in the axolotl oocyte will be investigated.

The basic steps involved in the molecular cloning of DNA are summarised in Old and Primrose (1981) where the advantages and disadvantages of the various host-vector systems are discussed in relation to the requirements of the project.

The ability to produce, by molecular cloning, large amounts of individual sequences overcomes many of the problems inherent in a study of repeated sequences. The major problem of alternative approaches is that whether genomic sub-populations are isolated by repeated buoyant density centrifugation (e.g. the global approach of Bernardi et al, (Cuny et al, 1981; Soriano et al, 1981)) or by restricted Cot histories (e.g. MacGregor and Andrews, 1977) the populations can never be regarded as homogeneous. Further, when reassociated products are used other technical problems are encountered. For example, the DNA is usually sheared so that the effective duplex length is quite small (see Appendix II); mismatching within the duplex regions means that subsequent manipulations e.g. nick translation, S_1 nuclease treatment to remove single strand tails may reduce the duplex length even further or alter the DNA sequence (see for example, Posakony et al, 1981). Molecular cloning provides a means of isolating individual stretches of DNA, in the native state, in relatively large quantities for study. Heterogeneous probes such as those described above are of course useful as guides to the choice and analysis of recombinant clones.

For the study of repeated sequence arrangement two basic approaches can be taken : (a) to purify long stretches of contiguous chromosomal sequence and investigate the repetitive and non-repetitive sequence relationships within that stretch of DNA (see for example Wensink et al, 1978; Musti et al, 1981); or (b) to analyse short range relationships and sequence family organisation across the genome. This would be more useful in an analysis of the evolutionary flexibility of the genome (e.g. Anderson et al, 1981; Scheller et al, 1978, 1981; Klein et al , 1978; Thomas et al, 1978; Moore et al, 1978, 1981).

Although the second option can be achieved via the first it was decided to study the genomic organisation of repeat sequences by isolating relatively short genomic sequences. This is most conveniently achieved in a plasmid vector system.

When the work was undertaken the guidelines for recombinant DNA research issued by the Genetic Manipulation Advisory Group (GMAG) were under review. Consequently work on amphibian genomic DNA was reclassified from that defined by the Williams Report (Cmnd 6600, 1976, Section 2.4, para. iii). Thus initial work using the crippled host strains E. coli χ1776 and MRC2 was superceded by work using the E. coli K12 HB101 - pAT153 host vector system. The clones described in this section were formed using this system. Fig. 5.1 summarises the cloning strategy adopted.

The choice of restriction endonuclease for the cloning step was determined by several factors. Two major factors involved were first, the presence of suitable vectors for the insertion of restriction fragments, such that recombinants could be selected. Second, the size distribution of the genomic restriction fragments produced after suitable endonuclease digestion.

Plasmid pAT153 is a high copy number derivative of pBR322 (Bolivar et al , 1977; Twigg-Sherratt, 1978). The plasmid has four unique

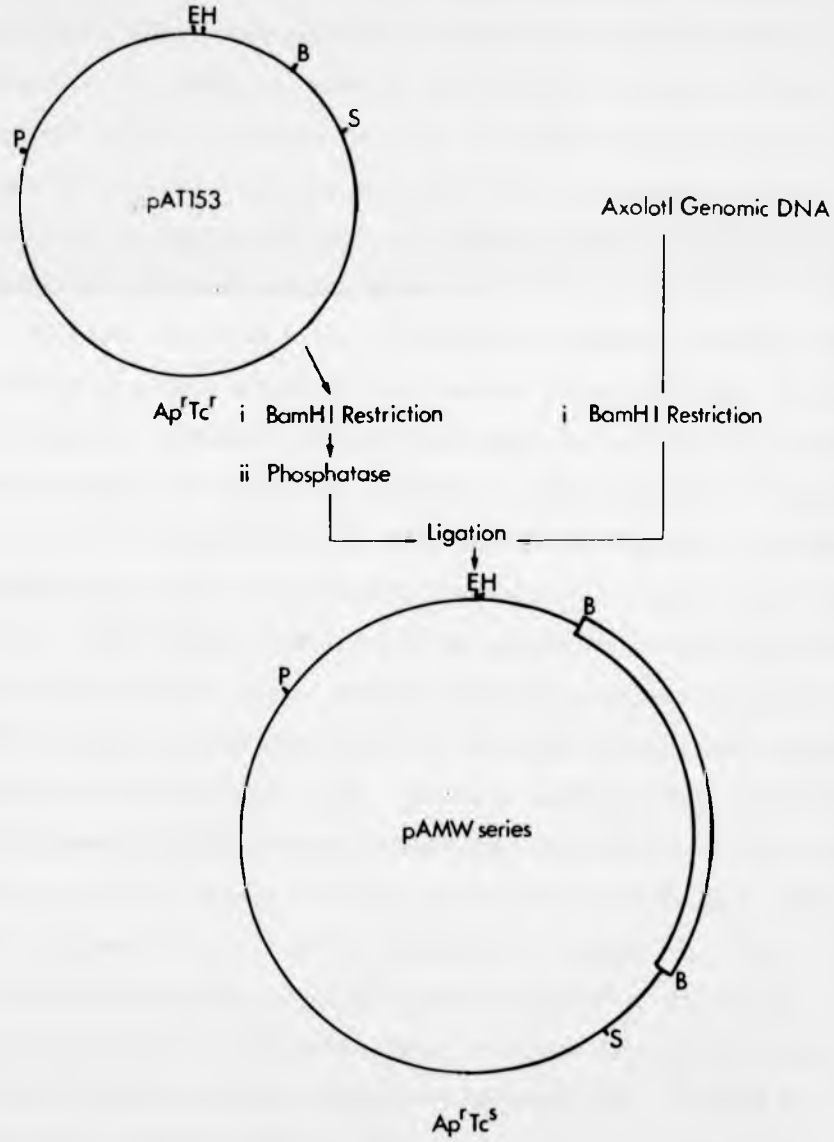
LEGEND TO FIG. 5.1

SCHEMATIC FOR "SHOTGUN" CLONING OF AXOLOTL DNA

Diagram describing the steps involved in "shotgun" cloning of axolotl genomic BamHI generated restriction fragments.

Plasmid pAT153 DNA is linearised with BamHI and dephosphorylated with bacterial alkaline phosphatase. Axolotl genomic DNA is restricted with BamHI and the DNAs are mixed and ligated. Recombinants are transformed into $CaCl_2$ treated E. coli HB101 and are selected on the basis of the change in phenotype from Ap^rT^r to Ap^rTc^s.

P	PstI
E	EcoRl
H	HindIII
B	BamHI
S	SalI
Ap	Ampicillin
Tc	Tetracycline
r	resistance
s	sensitivity


EH
B
S
P
pAT153
Axolotl Genomic DNA
Ap^r Tc^r
i BamH I Restriction
i BamH I Restriction
ii Phosphatase
Ligation
EH
B
P
pAMW series
B
S
Ap^r Tc^s

restriction sites within the two antibiotic resistance markers located on the plasmid; HindIII, BamHI, SalI within the tetracyline resistance genes and PstI within the ampicillin resistance gene as well as several other unique sites located round the plasmid e.g. EcoRl, AvaI, PvUII (Sutcliffe, 1978). However both the HindIII and PstI sites have been designated as "expressing" sites by GMAG, that is, insertion of DNA fragments into these sites might result in expression of the inserted fragments. Therefore cloning into these sites requires that the experiments be done at one level of containment higher than that for cloning into the BamHI or SalI sites.

When the size distribution of restriction fragments generated after digestion of axolotl genomic DNA was examined it was noted that the size distribution of fragments produced using BamHI, PstI or HindIII were very similar (Fig. 5.2A, lanes a, b, d) forming a continuous smear of fragment sizes with some indication of discrete bands within the smear. The SalI fragments were however predominantly high molecu lar weight, (Fig. 5.2A, lane 1). This digestion pattern has been noticed on numerous occasions and can be concluded to be a terminal digestion pattern as supercoiled BR322 included as internal control is completely digested under identical digestion conditions (Fig. 5.2B). The SalI digestion pattern is probably a consequence of methylation of genomic DNA within the restriction enzyme sequence, at the internal C G dinucleotide (Roberts, 1978; Bird, 1980). (See also lanes 5.2 g - i and the discussion in Chapter 3). It is therefore probable that, were this enzyme to be used in the cloning procedure, the genomic fragments cloned would have internal SalI sites as these would not be methylated in the bacterial host. It would be difficult to determine whether such sequences were in fact contiguous in the genome.

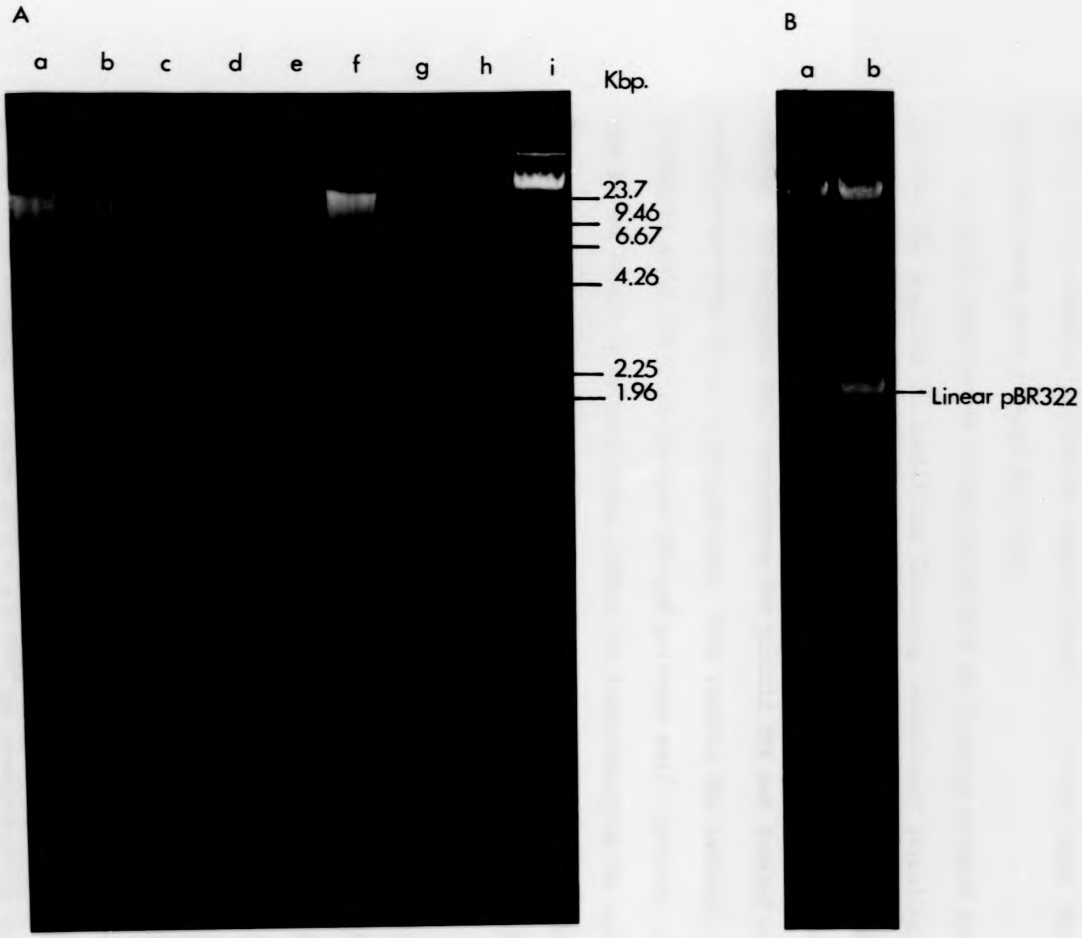
The enzyme BamHI therefore provides a suitable enzyme for cloning axolotl genomic fragments. When axolotl DNA is digested with this enzyme

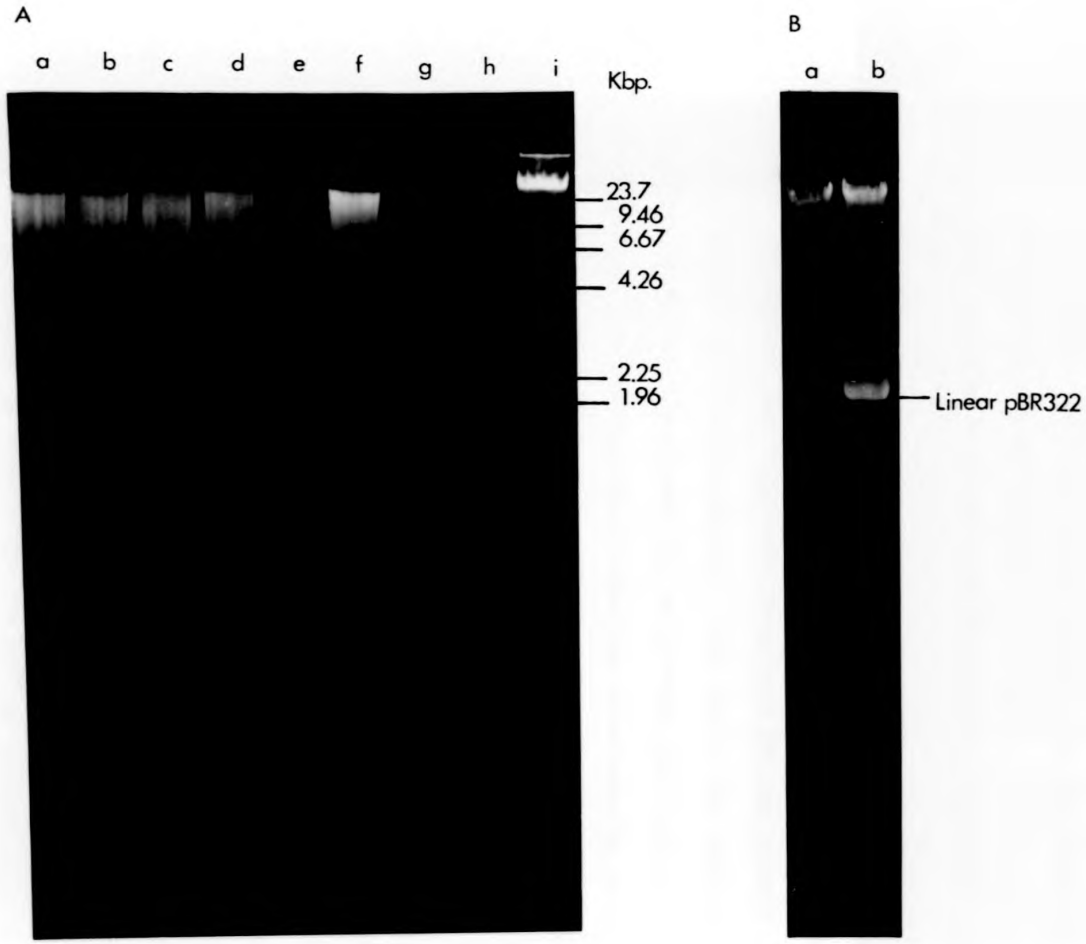
LEGEND TO FIG. 5.2

CHOICE OF RESTRICTION ENZYMES FOR CLONING

a. 2 μg of axolotl DNA was digested to completion with the relevant enzymes by incubating with a five-fold excess of enzyme for five hours using the manufacturers recommended conditions of assay. Lane a, BamHI; b, Pst I; c, EcoRl; d, HindIII; f, Hpa I; g, Msp I; h, Taq I; i, SalI; lane e shows DNA digested with HindIII. Fragment sizes are shown in kb.

b. 1 μg of axolotl DNA was digested with SalI as above either without (a) or with (b) 1 μg of supercoiled plasmid pBR322 (Bolivar et al. 1978). Marker DNA (3) was λ-HindIII as above.


A
a b c d e f g h i
Kbp.
23.7
9.46
6.67
4.26
2.25
1.96
B
a b
Linear pBR322


A
a b c d e f g h i
Kbp.
23.7
9.46
6.67
4.26
2.25
1.96
B
a b
Linear pBR322

a series of bands of varying molecular weights can be seen above the background smear when the gel slot is overloaded (5 - 10 μg of genomic DNA) (not shown) indicating a regular arrangement of BamHI sites within repetitive DNA, with either a regularly repeated distance between BamHI sites in tandemly repeated sequences, or the presence of two spaced BamHI sites within a particular repeat element. In some cases the observed bands were several kbp long.

In early experiments it was noted that by ligating plasmid and genomic DNA together in conditions favouring recombinant formation (Dugaizyk et al, 1975; Graf et al, 1979) in the region of 80% of the recombinanats were Tc^R indicating religation of the vector (data not shown). To minimise this occurrence the plasmid DNA was treated after restriction with alkaline phosphatase. This removes the terminal 5' - phosphate group from the plasmid DNA and prevents self ligation. As only one strand needs to be covalently linked for transformation the genomic DNA can provide the necessary 5' - phosphate group. This step reduces the background of religated plasmids and so simplifies the screening process.

Axolotl DNA was digested to completion with BamHI and was mixed with BamHI digested, phosphatase treated pAT153. The mass ratio of axolotl DNA to plasmid DNA was 1:1 at a total DNA concentration of 50 μg/ml to optimise circle formation (Dugaizyk et al, 1975). After 48 hours at 4°C an aliquot was removed for transformation. The remainder was stored at -20°C.

Transformation into $CaCl_2$ treated E. coli HB101 was as described in the Methods section. Using pure plasmid DNA transformation rates of 2 - 5 x 10^5 transformants/ μg were usually achieved with E. coli HB101. (This can be compared to rates of 2 - 4 x 10^3/ μg produced using the crippled host E. coli Y1776 (data not shown)). However the observed rate of transformation of the ligation mix was less than 10^4/ μg. This is presumably due to the presence of small DNA fragments in the mixture which

compete for the available binding sites on the surface of the $CaCl_2$ treated bacteria.

The ampicillin resistant transformants were picked out onto plates containing both ampicillin and tetracycline to establish which clones were tetracycline sensitive, indicating an insertion at the BamHI site. In this particular transformation 86% were ampicillin resistant and tetracycline sensitive. Around 200 colonies were picked onto nitrocellulose filters on nutrient plates containing ampicillin. E. coli HB101, transformed with pAT153, were also picked as controls. The colonies were allowed to grow overnight. The filters were then transferred to plates containing both ampicillin to maintain selection, and chloramphenicol, to amplify the plasmids (chloramphenicol prevents host protein synthesis but allows DNA replication to occur (Hershfield et al, 1974) . Experiments in which the chloramphenicol step was omitted resulted in very poor signal response with most colonies showing no hybridisation, even those subsequently shown to contain recombinant clones (see also Grunstein and Wallis, 1979)).

The colonies were lysed in situ as described by Humphries et al (1978) with the modifications noted in the Methods section. The filters were hybridised with axolotl DNA radioactively labelled with P^{32} by "nick translation" (Rigby et al, 1977). The DNA used for the probe was prepared by reassociating sheared (weight average 500 bases) axolotl genomic DNA to Cot50. The DNA was then treated with S_1 nuclease to remove single strand tails, to reduce possible contamination of the probe by adjacent sequences. The DNA was hybridised as described in the Methods section. Under the conditions of hybridisation employed it was expected that clones containing repetitive sequences would be clearly visible after the filters were washed and autoradiographed. Fig. 5.3 shows the results of this screening process. One clone, reference no. 1184, hybridised very strongly in this experiment. The area including the DNA print of clone 1184 was cut from the nitrocellulose so that subsequent exposures could be made to

LEGEND TO FIG. 5.3

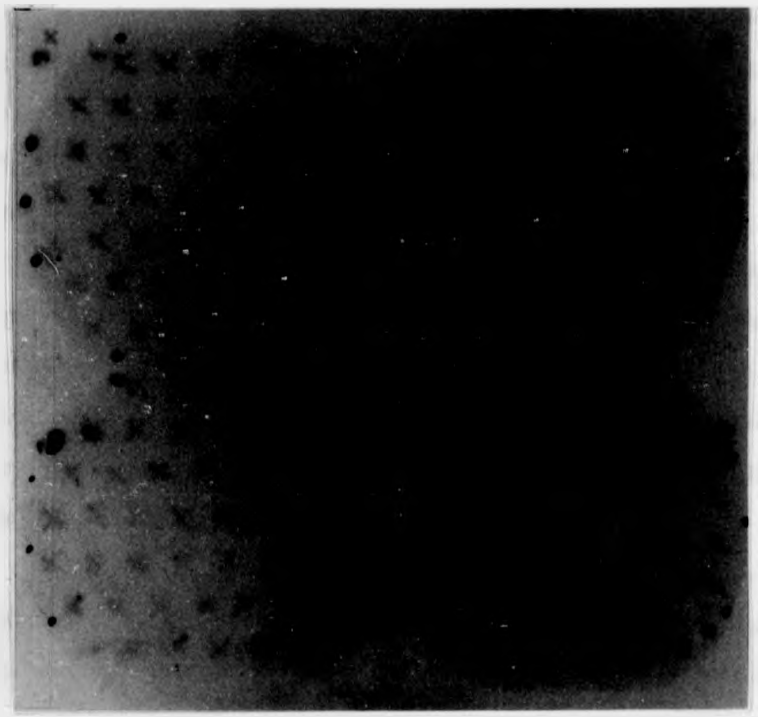
SCREENING OF RECOMBINANT CLONES WITH

COT 0 - 50 DUPLEX DNA

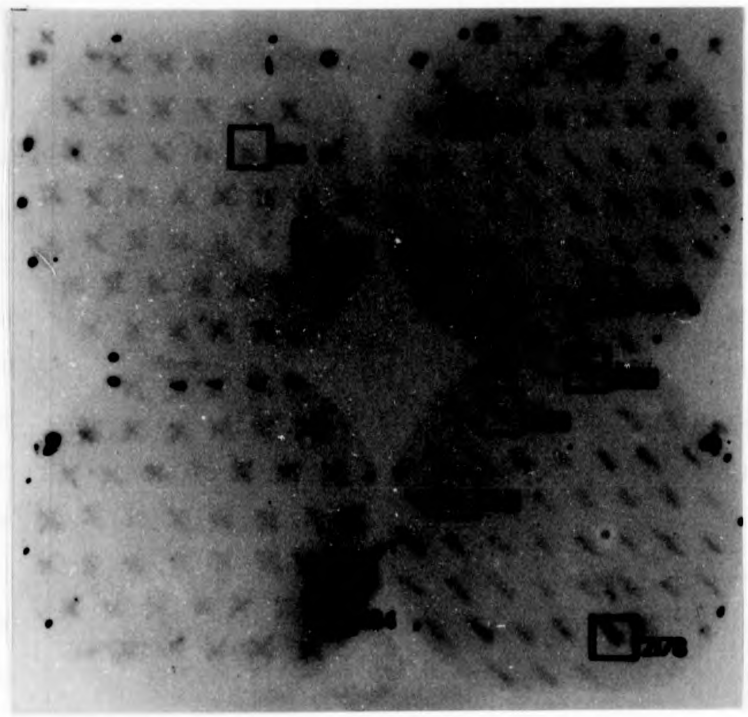
Around 200 $Ap^{R}Tc^{S}$ recombinants were streaked onto a nitrocellulose filter on the surface of a nutrient plate containing 100 μg/ml ampicillin. The colonies were grown overnight. The filters were lifted onto further plates containing 100 μg/ml ampicillin and 200 μg/ml chloramphenicol (see text). The filters were processed as Methods and hybridised with Cot 0 - 50 duplex DNA. Approximately 5 x 10^{6} cpm (10^{7} cpm/μg) were applied in a total of 5 mls of 3 x SSC/0.1% SDS/1 x Denhardts per filter and incubated at 65°C overnight. The filters were washed extensively in 3 x SSC at 65°C. The filters were exposed at -70°C with an intensifying screen 16 hours at -70°C.

(◂) E. coli HB101/pBR322 control colonies

The numbers refer to particular clones referred to in some detail in the text.



Probe Cot 0-50



Probe Cot 0-50

pick up less intensely hybridising clones. Negative controls on these filters (*E. coli* HB101 containing pAT153) are marked (◀). As the probe is heterogeneous in terms of sequence the hybridisation signal may reflect the frequency of the cloned sequence within the probe, or the presence of tandem repeated elements within the recombinant plasmid, or both. It was decided to isolate clones exhibiting a range of signal response in order to obtain a range of the possibilities noted above. The transformants were also hybridised with probe corresponding to reassociated Cot5 duplex. A subjective estimate of the response to both hybridisation probes of several clones which were subsequently analysed is shown in Table 5.1.

Recombinant plasmid DNA was isolated from several clones. Initially small scale preps were used to determine the presence of a plasmid within the bacteria and the size of the plasmid to give some indication of the insert size. Larger scale preparations were prepared for several recombinants and the plasmid DNA was digested with BamHI to determine the size of the cloned genomic DNA (Fig. 5.4).

The figure shows the digestion products of several tetracycline sensitive clones separated by size on agarose gels. It is apparent that in many cases the insert is not released by the BamHI digestion, even though the plasmids are linearised, suggesting the presence of only one BamHI site.

It has been suggested that damage may have occurred to free BamHI ends during the ligation process as similar problems were experienced by others using the same batch of DNA ligase (P. C. Turner, pers. comm.).

Clones pAMW2135 and pAMW2138 produce fragments suggesting the presence of two genomic fragments, one of which is released by BamHI digestion while one remains contiguous with the pAT153. As there is no evidence to suspect that the genomic restriction was only partial, these fragments presumably originated from distinct regions in the genome.

TABLE 5.1

RESPONSE OF CLONES TO GENOMIC PROBES

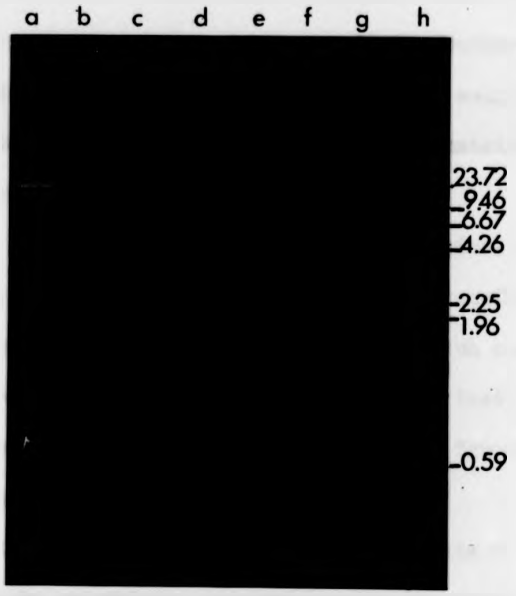
CLONE	PROBE	
	Cot 0 - 5	Cot 0 - 50
1114	-	+/-
1131	-	+/-
1184	+++	++++
1199	+++	+++
2126	-	+/-
2135	+	+
2138	+	+/-
2151	-	-
2178	-	+
pAT153	-	-

LEGEND TO FIG. 5.4

BamHI DIGESTION OF PUTATIVE AXOLOTL RECOMBINANT PLASMIDS

Plasmid DNA was isolated from a number of putative axolotl clones which gave a positive signal in a Grunstein-Hogness hybridisation. The DNA was digested with BamHI to release the inserted axolotl DNA. The DNA was separated by electrophoresis through 0.8% agarose. Size markers were λ-HindIII restriction fragments (lanes a and h).

In all cases except lane e (1199) the vector and insert DNAs are not separated by BamHI. In lane e the vector DNA corresponds to the lower, more rapidly eletrophoresing band.



a,h	λ -Hind III
b	pAMW1114
c	1131
d	1184
e	1199
f	2135
g	2138

A basic restriction map was produced for several of these clones using common restriction enzymes which cleave at hexanucleotide recognition sequences. These are summarised on Fig. 5.5, which also indicates the estimated size of the insert. It is clear that a wide range of sizes of insert are represented. Clones 2126, 2151 and 2178 however gave anomalous results. From the colony hybridisation it appeared that these clones contained axolotl sequences. However restriction enzyme mapping as well as Southern blot hybridisation of these clones to genomic DNA (not shown) suggest that the insert was lost at some stage, or that the original Grunstein-Hogness hybridisation patterns were anomalous.

Clones 2135 and 2138 produced identical restriction maps suggesting that these clones arose from a single transformation event. The presence in both clones of two inserts one of which can be removed by BamHI and one which cannot, also tends to suggest that these are "daughter" clones formed after a cell division during the "growing out" period in the transformation procedure (see Methods).

Fig. 5.6A-C shows restriction digests of the clones using restriction enzymes recognising tetranucleotide sequences. Clones 2126, 2151 and 2178 were included to determine whether very small inserts were present within the clones. The digests of 2126, and 2151 suggest that no insert is present. 2178 however gives anomalous results with several enzymes. This suggests that some form of internal rearrangement may have occurred, although no net increase in size can be shown which may indicate the presence of an insert. Clone 2138 is not shown in A-C as the restriction pattern was found to be identical to 2135 for all the enzymes used. The restriction analysis of the clones indicated a number of features of interest. First the heterogeneity in restriction pattern (with the exception of 2135 and 2138) confirmed that the clones contained independent genomic sequences. Secondly, for clones 1114, 1199 and 2135 the abundance of sites detected

LEGEND TO FIG. 5.5

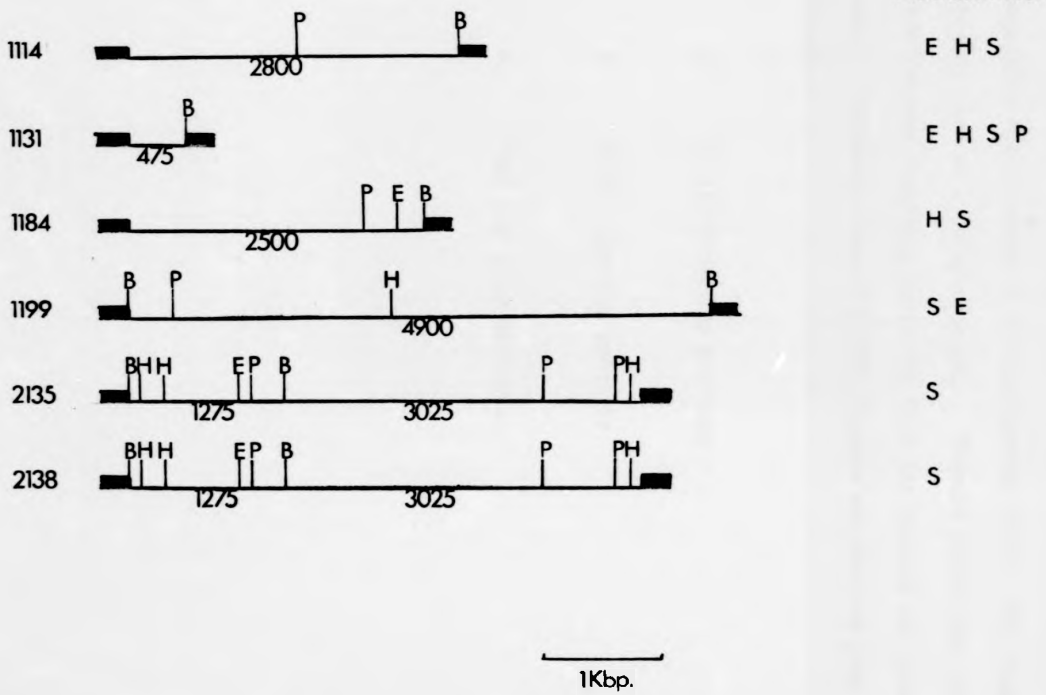
CURRENT RESTRICTION MAPS OF AXOLOTL RECOMBINANT CLONES

Selected axolotl recombinant clones were digested with various enzymes and restriction maps were generated in the manner described in Appendix III. Those enzymes which were tested which do not cleave within the insert are noted. The apparent size in bp of the insert has been noted below the relevant sequence.

The filled in regions represent pAT153 sequence.

B - BamHI; P - Pst I; E - EcoRl; H - HindIII; S - SalI.

In some cases the location of other sites has been deduced e.g. from Fig. 5.6. However in most cases these sites have not been confirmed by detailed mapping. Partial digest mapping of clone 1199 has been performed and relative locations of restriction sites have been determined for a number of enzymes. These partial and incomplete data are shown in Appendix III (Fig. AIII, 2).


No sites in insert
1114
P
B
2800
E H S
1131
B
475
E H S P
1184
P E B
2500
H S
1199
B P H B
4900
S E
2135
B H H E P B P P H
1275
3025
S
2138
B H H E P B P P H
1275
3025
S
1Kbp.

LEGEND TO FIG. 5.6 A - C

RESTRICTION ENZYME DIGESTS OF SELECTED AXOLOTL CLONES

Selected axolotl recombinant clones were digested with several enzymes recognising and cleaving at tetranucleotide sites. The fragments were separated by size on 2.5% agarose gels. Plasmid pAT153 was included to distinguish between fragments generated from the plasmid and those from the insert. Fragment sizes of pAT153 digests are derived from Sutcliffe (1978).

A. HaeIII digestion patterns

B. HinfI digestion patterns

C. TaqI digestion patterns.

LEGEND TO FIG. 5.6 A - C

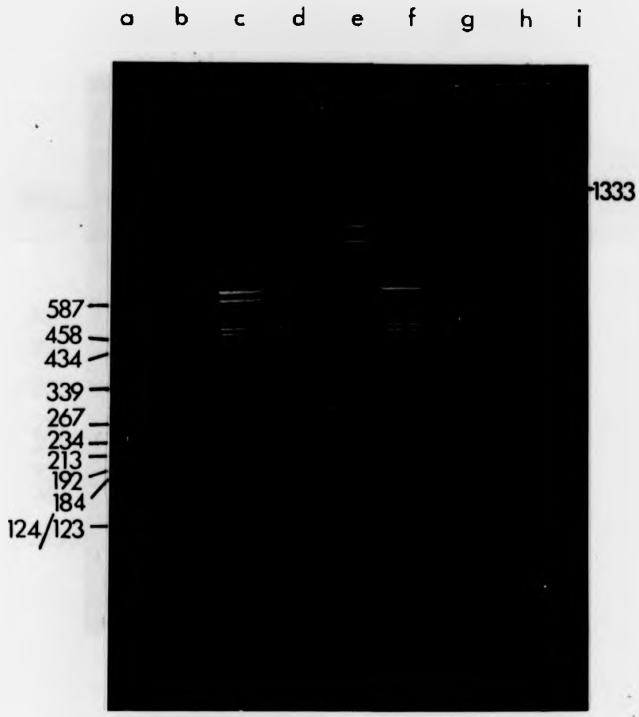
RESTRICTION ENZYME DIGESTS OF SELECTED AXOLOTL CLONES

Selected axolotl recombinant clones were digested with several enzymes recognising and cleaving at tetranucleotide sites. The fragments were separated by size on 2.5% agarose gels. Plasmid pAT153 was included to distinguish between fragments generated from the plasmid and those from the insert. Fragment sizes of pAT153 digests are derived from Sutcliffe (1978).

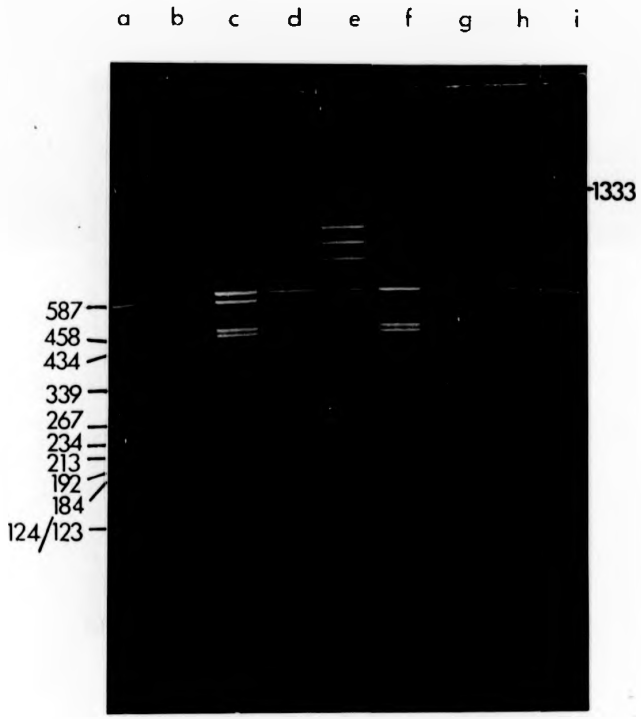
A. HaeIII digestion patterns

B. HinfI digestion patterns

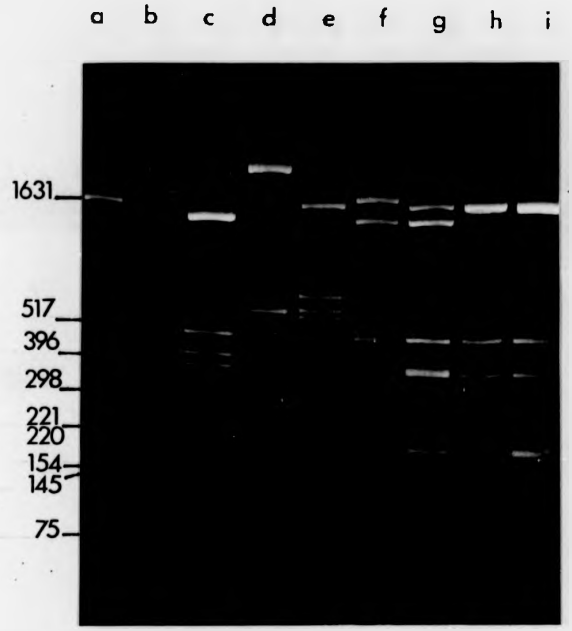
C. TaqI digestion patterns.



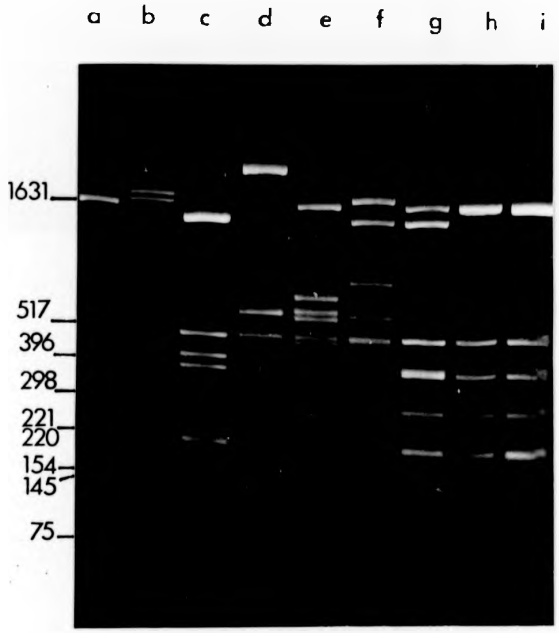
a	pAT153
b	pAMW1114
c	1131
d	1184
e	1199
f	2126
g	2135
h	2151
i	2178



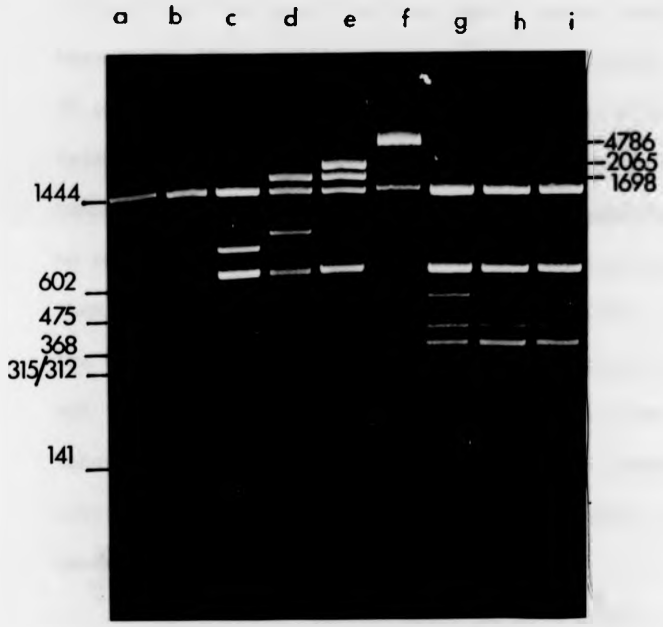
a pAT153
b pAMW1114
c 1131
d 1184
e 1199
f 2126
g 2135
h 2151
i 2178



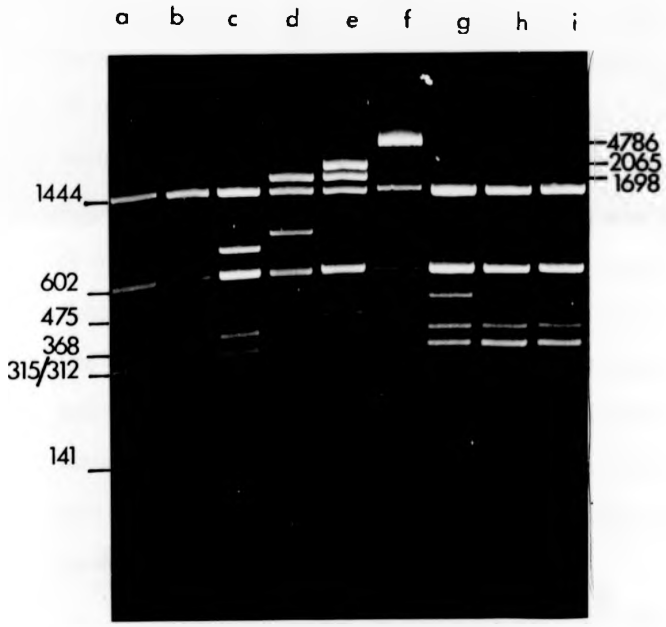
a pAT153
b pAMW1114
c 1131
d 1184
e 1199
f 2135
g 2178
h 2126
i 2151



a pAT153
b pAMW1114
c 1131
d 1184
e 1199
f 2135
g 2178
h 2126
i 2151



a	pAT153
b	pAMW1114
c	1131
d	1184
e	1199
f	2135
g	2178
h	2126
i	2151



a	pAT153
b	pAMW1114
c	1131
d	1184
e	1199
f	2135
g	2178
h	2126
i	2151

(Figs. 5.5, 5.6) is sufficient to rule out the possibility of internal reiteration. For 1131, the number of restriction sites within the insert is too low to rule out completely the faint possibility of internal reiteration. Clone 1184 however clearly indicates the presence of internal repetition. There appears to be a regular arrangement of HaeIII sites within the insert with a common repeat length of approximately 57 base pairs (Fig. 5.6A). The observations suggest that there may be up to 36 copies of the monomer sequence as well as a trimer (171 bp) and a tetramer (228 bp) sequence. This tandem arrangement would account for the presence of few sites for the other tetranucleotide cutters. If there were no sites for these enzymes in the original sequence before amplification then sites would only be introduced by mutation. This would account for the internal Pst and EcoRl sites, which occur only once in the insert, and the occasional loss of an HaeIII site to form the multimers observed. Taken together with the colony hybridisation results the restriction pattern suggests that the sequence may be quite highly repetitive in the genome.

It was decided that it would be more useful to study in some detail a small number of clones, each chosen for some particular feature of interest. Therefore the most complete data exists for three of the clones so far described. These are 1131, chosen because it has a small insert, 4 75 bp, of low copy number in the genome; 1184 (2.5 kb) because it appears to have a short tandemly repeated sequence within the insert; and 1199 because the large (4.9 kb) insert may represent a complex repeat organisation. A less complete analysis was also performed on various other clones. The repetition frequency of the cloned sequences in the genome was determined by two methods. First, by reassociation kinetics in the presence of sheared genomic DNA and secondly, by saturation hybridisation against genomic DNA. Other factors which affect the rate of reassociation such as % G+C and family divergence (see Appendix II) were also investigated to

attempt to produce a more accurate estimation of the genomic representation of these clones.

The kinetic experiments were of two kinds. First, the clones were allowed to self reassociate to determine whether any sequences were repeated within the insert. Second, the clones were used as tracers in reassociations driven by a vast excess of cold sheared axolotl DNA. For this purpose the whole recombinant plasmid was radioactively labelled with ^{32}P by "nick translation". This reduces the DNA to an average single strand size of around 400 bases. The driver DNA for the tracer-driver reassociation experiments had an average single strand size of 500 bases. The difference in lengths of tracer and driver does not warrant corrections for length effects (see Chamberlin et al, 1978). Plasmid pAT153 and E. coli DNA were also reassociated, to provide single copy controls.

Fig. 5.7 shows the self reassociation kinetics of the various DNA's, while the various kinetic estimates are shown in Table 5.2.

Clones 1184 and 1199 exhibit a small fraction of the clone which appears S_1 nuclease resistant at early Cot points, even after normalisation has been performed. This may also be a labelling artefact, but may be due to the presence of inverted repeats within the clone. Clone 1184 also appears to have a high degree of internal repetition. The calculated reiteration frequency of 47.5 tends to agree with that estimated from the restriction analysis (Fig. 5.6). Only 30% of the clone reassociates with kinetics faster than that for a single copy sequence, although the insert represents 40% of the clone. This discrepancy could be explained as an effect of the tandemly repeated nature of the insert. Therefore the low sequence complexity of 36bp (instead of 57bp) reflects this discrepancy.

When the nick-translated clones were used as tracers in reassociations driven by an excess of sheared genomic DNA the moves shown in Fig. 5.8 were obtained. The kinetic parameters derived from these curves are given as

LEGEND TO FIG. 5.7

PLASMID SELF REASSOCIATION

Plasmid DNA was labelled with ^{32}P by nick translation. Small quantities were mixed with cold, sheared plasmid DNA. ^{3}H E. coli DNA was added to each reassociation mix as an internal standard for the reassociation. After reassociation to various Cot points aliquots were removed for S_1 nuclease analysis.

After plotting the data several adjustments were made as follows. Possibly due to the nick translation procedure all of the plasmid curves showed a fraction of the counts as double stranded at the earliest Cot points. These have been normalised to 100% based on the initial value for the vector pAT153 which was 11% in this experiment. The K for E.coli DNA in the pAT153 reassociation was observed to 0.523. The values for the E. coli DNA reassociation rate in the other reassociations was adjusted to this value and the observed curves for the plasmids were adjusted by the same amount for comparison with pAT153. Kinetic estimates based on this figure are shown in Table 5.2.

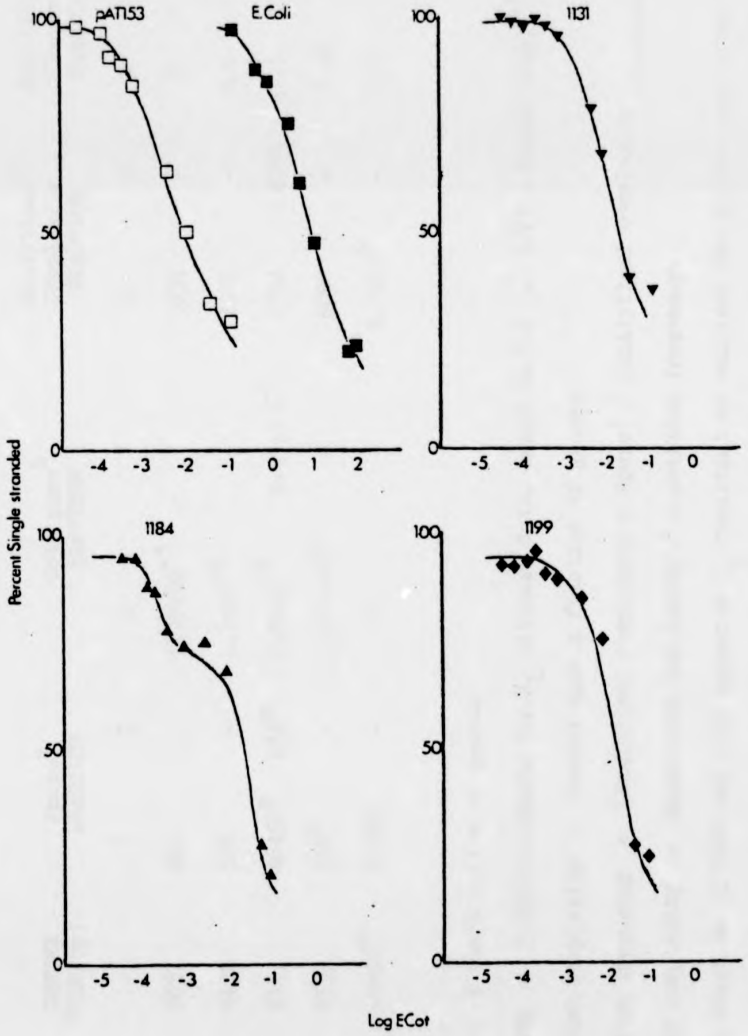

pAT153
E.Coli
1131
1184
1199
100
50
0
-4 -3 -2 -1 0 1 2
-5 -4 -3 -2 -1 0
Percent Single stranded
Log ECot

TABLE 5.2 PLASMID SELF REASSOCIATION

DNA	GENOME SIZE(bp)	Kpure[a] FRACTION		$Cot_{\frac{1}{2}}$ pure[b] FRACTION		Analytical[c] Complexity FRACTION		Repetition[d] Frequency FRACTION		Sequence[e] Complexity FRACTION	
		1	2	1	2	1	2	1	2	1	2
pAT153	3657	461	-	7.95×10^{-3}	-	3657	-	1	-	3657	-
pAMW1131	4132	326	-	1.1×10^{-2}	-	4132	-	0.8	-	5171	-
pAMW1184[f]	6157	46457[g]	434[g]	7.9×10^{-5}	8.45×10^{-3}	1724	4186	47.5	1.08	36	3875
pAMW1199[f]	8554	306[h]	-	1.197×10^{-2}	-	8088	-	1.47	-	5502	-
E.coli	4.5×10^{6}	0.523	-	7	-	4.5×10^{6}	-	1.4	-	3.2×10^{6}	-

NOTES

a. Kpure = K obs/fraction of genome

b. $Cot_{\frac{1}{2}}$ pure = 3.6661161/Kpure for S_1 nuclease data, based on 0.5 = $1/(1 + \text{Kpure } Cot_{\frac{1}{2}})^{0.45}$

c. Analytical complexity = Genome size x fraction of genome

d. Repetition frequency = (Analytical complexity x Kpure) / (Analytical complexity $_{\text{standard}}$ x Kpure $_{\text{standard}}$)

e. Sequence complexity = Analytical complexity / repetition frequency.

f. A small fraction of 1184 and 1199 appeared S_1 resistant at earliest Cot points even after normalisation.

g. Kpure's for 1184 were calculated based on two second order components with K obs of 13008 and 291 and comprising 0.28 and 0.67 of the plasmid sequence.

h. Kpure for 1199 was calculated based on one second order component with K obs of 291 and comprising 0.95 of the plasmid sequence.

LEGEND TO FIG. 5.8

DRIVER-TRACER REASSOCIATIONS

Plasmid DNA was labelled by nick translation and reassociated in the presence of a vast excess of cold, sheared axolotl genomic DNA. For most of the data points a driver ratio of 10^5 : 1 was used. For some of the early points a 10^4 : 1 ratio was used. At various times aliquots were removed for assay. In each case the data has been normalised as for Fig. 5.7 but no rate adjustments have been made. A single copy rate of 0.523 for E. coli (4.5×10^6 bp haploid genome) (Fig. 3.9) has been used to derive rate constants for the data described here.

The x axis refers to driver Log E Cot.

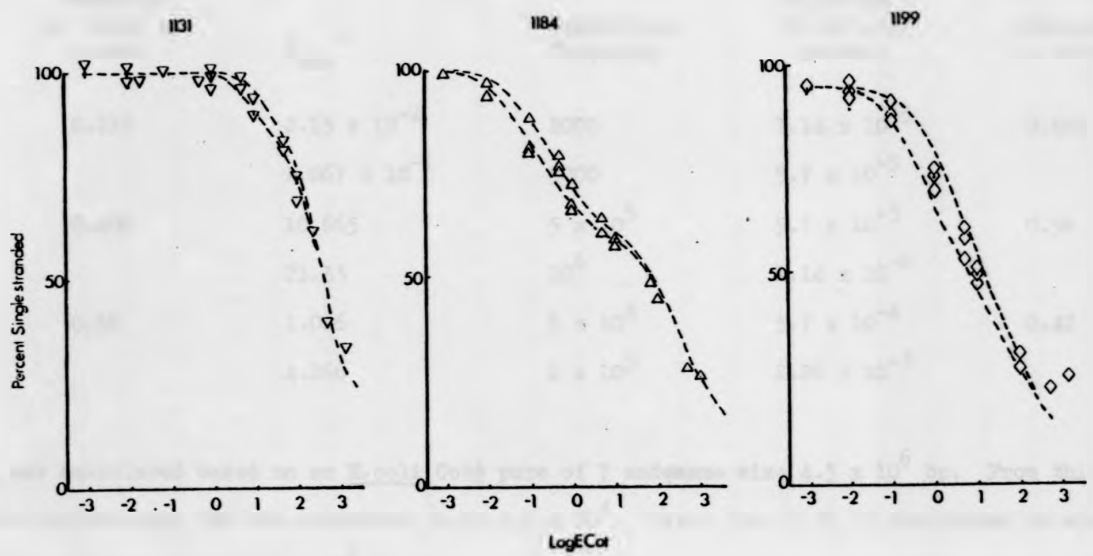

1131
1184
1199
100
50
0
Percent Single stranded
-3
-2
-1
0
1
2
3
LogECat

TABLE 5.3 DRIVER-TRACER REASSOCIATIONS

CLONE	Fraction of clone as insert	K_{obs} [a]	Repetition frequency	Fraction [b] of axolotl genome	Fraction as vector	K_{obs} [a]
1131	0.115	2.13×10^{-2}	1000	1.14×10^{-5}	0.885	750
		1.067×10^{-1}	5000	5.7×10^{-5}		
1184	0.406	10.665	5×10^{5}	5.7×10^{-3}	0.58	795
		21.33	10^{6}	1.14×10^{-2}		
1199	0.58	1.066	5×10^{4}	5.7×10^{-4}	0.42	680
		4.266	2×10^{5}	2.28×10^{-3}		

a. K_{obs} was calculated based on an E.coli $Cot\frac{1}{2}$ pure of 7 and genome size 4.5×10^{6} bp. From this a $Cot\frac{1}{2}$ pure for axolotl single copy DNA was estimated to be 5.5×10^{4}. Given that 0.32 of the genome is single copy $Cot\frac{1}{2}_{obs}$ (single copy) would be 1.72×10^{5} in S_1 assayed reassociation curves. From this $Cot\frac{1}{2}_{obs}$ for single copy DNA hypothetical curves of a given repetition frequency were generated.

b. $$\text{Fraction of axolotl genome} = \frac{\text{Repetition frequency} \times \text{fragment length (b)}}{\text{Genome size (b)}}$$

Fragment length was assumed to be 400 bases.

Table 5.3. In each case the vector component was assumed to be equivalent to the fraction of the plasmid calculated from the restriction enzyme data. Because of the degree of scatter of the data points it was decided to plot two possible curves indicating the range of repetition frequency consistent with the observed data. For each of the three clones examined the range of repetition frequency plotted most probably encompasses the true repetition frequency. Repetition frequency shown by 5.3 refers to the average length of sheared DNA used in the reassociation expt (3 - 400 bases). It is not believed that the data at present warrant a more accurate estimate.

The estimates derived by reassociation kinetics were then compared to estimates derived from saturation experiments. In these experiments a fixed amount of axolotl DNA was bound to nitrocellulose. Identical filters were hybridised with increasing amounts of radioactively labelled recombinant DNA, or the vector pAT153. The amount of label hybridised to the filter bound DNA after 24 hours was measured and used to estimate the apparent copy number of the hybridising sequences in the genome. The data is plotted as Fig. 5.9 and the apparent copy numbers are derived in Table 5.4. The values of copies bound per genome shown in Table 5.4 refer to the full insert. These values must be multiplied by any internal repetition to give a true estimate of the reiteration frequency within the genome. It should also be remembered that these saturation values are minimum estimates. The values for 1131 from both reassociation kinetics and saturation hybridisation are in reasonable agreement, giving an estimated reiteration frequency in the range $1 - 5 \times 10^3$ copies per haploid genome. For 1184 the already observed internal repetition together with the saturation hybridisation results produce a value for the 57bp repeat of 2.35×10^6 copies, which is in reasonable agreement with the kinetic analyses. For 1199 however the saturation data indicate a minimum of 1.3×10^4 copies of the 4.9kb insert, while the kinetic data suggest

LEGEND TO FIG. 5.9

SATURATION HYBRIDISATION

Sheared axolotl DNA was denatured in NaOH, neutralised with Hcl and loaded onto millipore filters (0.45 μm) prewashed in 6 x SSC. The filters were baked at 80°C for two hours. Probe DNA was denatured by boiling in 0.1 x SSC and adjusted to 4 x SSC 50% Formamide 1 x Denhardts. Various amounts of probe DNA were loaded onto duplicate filters. Hybridisation was at 37°C for 24 hours.

The filters were washed in 6 x SSC at room temperature, dried at 60°C for 30 minutes and counted in Toluene/PPO/POPOP.

For probe 1131 (▼) Millipore 25 mm diameter filters were used and 50 μg of sheared, denatured axolotl DNA was added. For 1184 (●), 1199 (■) and pAT153 (▲) HAWP1300 13 mm diameter filters were used and 500 ng of sheared, denatured axolotl DNA was loaded.

Each point is the average of three identical hybridisations. The values shown for pAT153 (▲) have been subtracted from the corresponding values for the other clones before the curves were plotted.

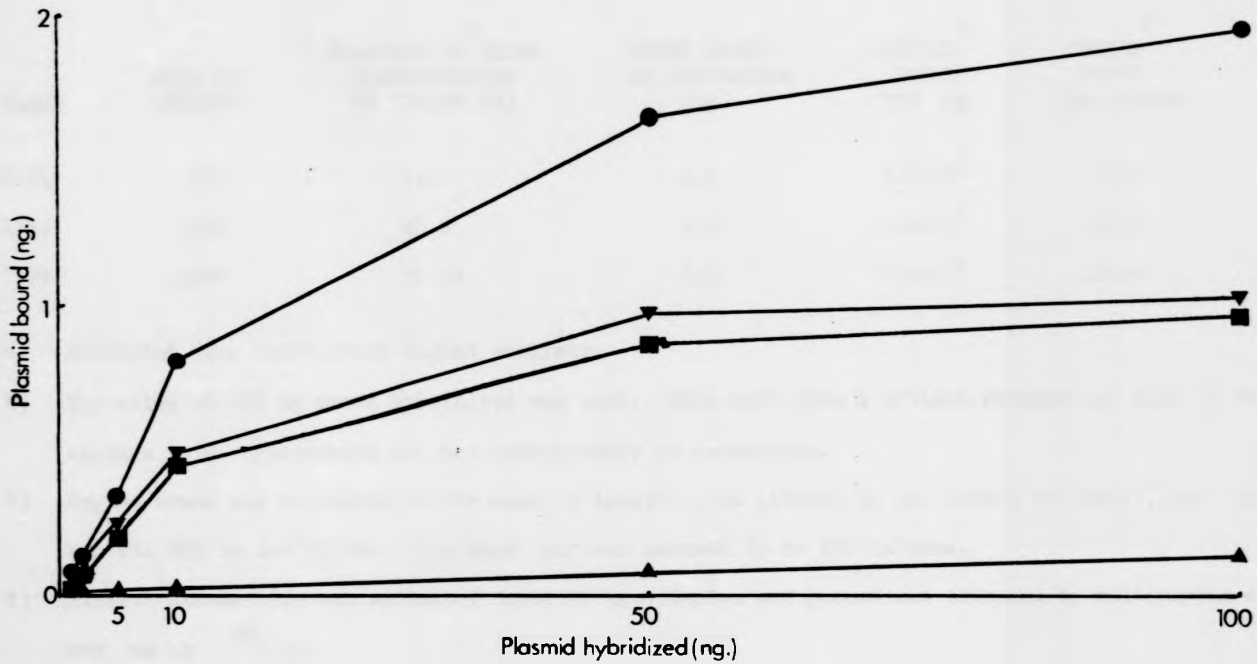

2
1
0
Plasmid bound (ng.)
5
10
50
100
Plasmid hybridized (ng.)

TABLE 5.4

ESTIMATE OF REITERATION FREQUENCY BY SATURATION HYBRIDISATION

CLONE	SIZE OF INSERTS	a FRACTION OF PROBE CORRESPONDING TO INSERT (%)	b PROBE BOUND AT SATURATION (ng)	c COPIES BOUND PER ug	d COPIES BOUND PER GENOME	e FRACTION OF AXOLOTL GENOME
1131	475	11.5	1.0	3.9×10^{7}	1481	2×10^{-5}
1184	2500	40.6	1.9	1.4×10^{9}	53485	3.8×10^{-3}
1199	4900	57.24	0.9	3.4×10^{8}	12926	1.8×10^{-3}

a) Estimated from restriction digest analysis.

b) The value at 100 ng probe hybridised was used. This will give a minimum estimate as each of the clones appears to be approaching but not demonstrably at saturation.

c) Copies bound was estimated as the mass of insert bound divided by the length of insert, per ug of axolotl DNA on the filter. One base pair was assumed to be 650 daltons.

d) Axolotl genome size was estimated to be 38 pg. Copies per genome was obtained by multiplying copies per ug by $^{38}/_{10^6}$.

e) Fraction of axolotl genome was calculated from $\frac{\text{Insert size x No. of copies}}{\text{genome size (bp)}}$

reiteration frequences in the range $5 \times 10^4 - 2 \times 10^5$ copies per genome. It is possible that within the 4.9 kb insert there exists separate repeat families with different copy numbers in the axolotl genome which produce the average value in the saturation experiment while in the driver-tracer reassociations the most highly repeated components are the most easily observed.

Several other factors influence the apparent repetition frequency measured by these methods. For example the familial divergence reduces the apparent rate of reassociation while the % G+C of the particular DNA sequence also has effects on rate of reassociation (see Appendix II). These two factors were investigated by controlled thermal denaturation of the clones before and after reassociation in the presence of axolotl genomic DNA.

In Section I of the Results the % G+C content of total axolotl DNA was calculated from denaturation experiments based on the formulae of Mandel and Marmur (1968). These formulae were applied to the Tm obtained by the controlled denaturation of the plasmid sequences from hydroxylapatite. Clone DNA, ^{32}P labelled and sheared by "nick translation" was mixed with approximately 50 µg cold, sheared E. coli DNA as internal standard. The DNA was loaded on hydroxylapatite at 60°C and subsequently eluted at 5°C increments with 0.12M sodium phosphate buffer. A final wash at 95°C with 0.4M sodium phosphate provided the 100% elution value. The values for the Tm of E. coli DNA were within 1°C of 90°C in all cases. Each set of curves has been adjusted so that the E. coli curve has a Tm of 90°C. Therefore the plasmid curves can be directly compared. Fig. 5.10 shows the melting curves generated for pAT153, 1131, 1184 and 1199. The Tm values and the % G+C values calculated from the data are given in Table 5.5 . It is clear from these results that the apparent % G+C for each clone is very different and varies from 35 to 52%. These values can be compared to the value of 46% obtained for total axolotl genomic DNA. Of

LEGEND TO FIG. 5.10

MELTING CURVES OF RECOMBINANT PLASMIDS

Plasmid DNAs were "nick translated" with P^{32}dNTP's and bound to hydroxylapatite in the presence of 50 μg sheared *E. coli* DNA as an internal standard for the denaturations.

The DNA was washed extensively at 60°C with 0.12M Na phosphate pH6.8 then the temperature was raised by 5°C increments. After 5 minutes at each new temperature the columns were washed with 2 x 3 mls 0.12M Na phosphate. At 95°C the columns were washed with 2 x 3 mls 0.12M Na phosphate, then 2 x 3 mls 0.4M Na phosphate. The columns were dissolved in 25% TCA and also counted to monitor irreversibly bound counts (Kiper and Herzfeld, 1978). These were usually less than 1% of the total and have not been considered further.

The fractions were monitored optically at 260 nm and by Cerenkov counting. Curves were plotted as the cumulative % counts released with increasing temperature. The curves were adjusted slightly to give a consensus Tm for *E. coli* of 90° C.

(a) pAT153

(b) pAMW1131

(c) pAMW1184

(d) pAMW1199

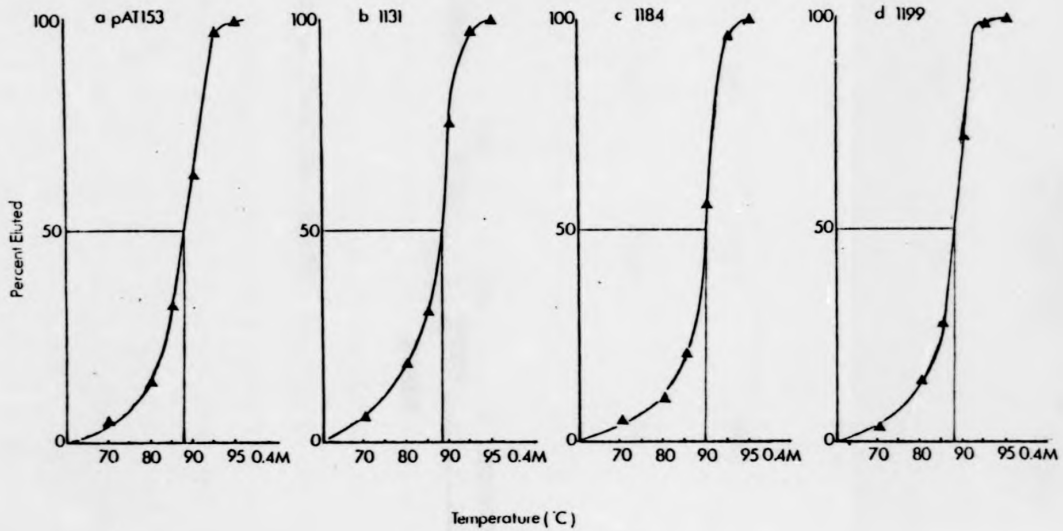

a pAT153
b 1131
c 1184
d 1199
100
50
0
Percent Eluted
70 80 90 95 0.4M
Temperature (°C)

TABLE 5.5

THERMAL ELUTION DATA

PLASMID	Tm	FRACTION OF PLASMID CORRESPONDING TO INSERT DNA	Tm [a] INSERT	%(G+C) [b] INSERT
pAT153	88	0	-	-
pAMW1131	87.5	0.115	83.6	35.3
pAMW1184	89	0.406	90.5	52.1
pAMW1199	88	0.572	88	46

a. $$Tm_{INSERT} = \frac{Tm_{TOTAL} - \left(Tm_{VECTOR} \times FRACTION_{VECTOR}\right)}{FRACTION_{INSERT}}$$

b. $\%G+C = 2.44\ (Tm - 69.3)$

the three clones mentioned here 1199 has a % G+C content most closely resembling bulk DNA.

In order to estimate the degree of divergence of each repeat family the nick translated clones were reassociated with genomic DNA to Cot 500 under conditions such that plasmid self reassociation would be minimised. A driver Cot of 500 was used to ensure that most family members would have the opportunity to reassociate. Control reassociations involving plasmid alone suggest that 10 - 15% of the apparent reassociation, under the conditions employed, may be due to plasmid self reassociation. The data have not been corrected for this. The observable effect would be to reduce the apparent degree of divergence, with greater errors occurring for more diverged families. However for the cases presented here the error would not exceed 1.5°C in the case of 1199 and would be less than 0.2°C for 1131 (assuming that the reassociated Tm for the plasmid is the same as the native Tm).

The results are shown as Fig. 5.11 and the calculated values for the divergence of each family are shown in Table 5.6. Fig 5.11 (bottom figures) also shows the derivate profile, that is the fraction of the DNA eluting at each temperature. From these graphs it can be seen that the elution profile for 1199 is observably different from that of the other two clones. A broader profile is seen which may suggest (by analogy to other systems e.g. Klein et al, 1978) that other members of the 1199 family exist which are unstable at the reassociation criterion which was used. The small amount of DNA eluted at 65°C (in all cases) may be due to the instability of hybrids with Tm close to the reassociation temperature (60°C). If this is so in the case of 1199 then the family may be much larger than predicted on the basis of saturation or kinetic analyses, which were also performed at the same criterion of reassociation.

The divergence values shown in Table 5.6 indicate that the family represented by clone 1131 is quite homogeneous. Taken together with the

LEGEND TO FIG. 5.11

ANALYSIS OF DIVERGENCE OF REPEAT FAMILIES

Recombinant clones were nick translated with ^{32}PdNTP's to high specific activity (5×10^7 cpm/ μg). Trace amounts were reassociated with a 10^5 mass excess of sheared genomic DNA to driver Cot 500. Aliquots were bound to hydroxylapatite and eluted as Fig. 3.12.

Top, Cumulative elution profile

Bottom, Fraction eluted at each temperature

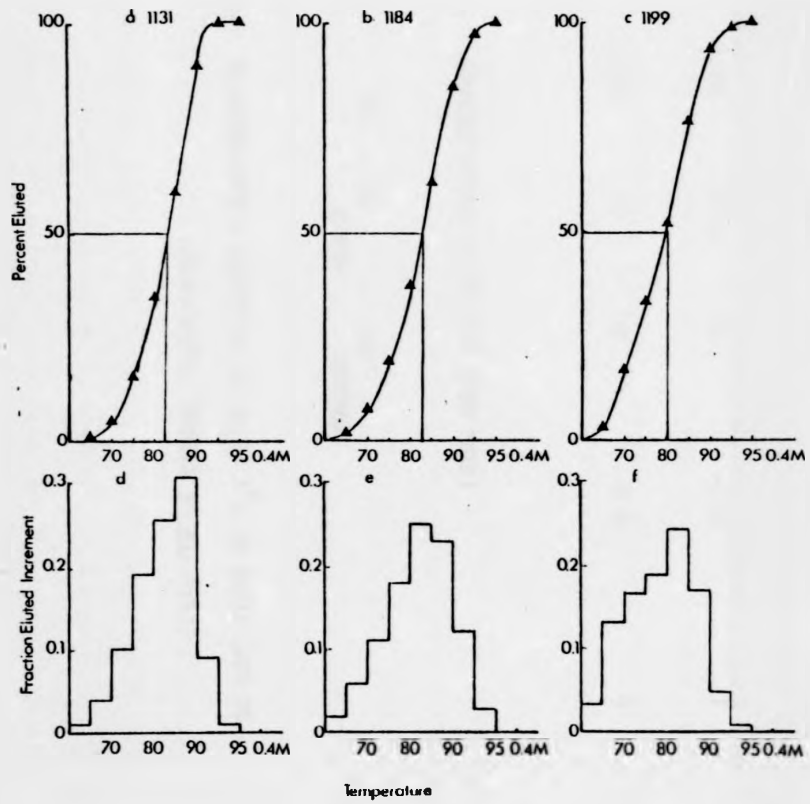

a 1131
b 1184
c 1199
100
50
0
Percent Eluted
70 80 90 95 0.4M
d
e
f
0.3
0.2
0.1
Fraction Eluted Increment
Temperature

TABLE 5.6

DIVERGENCE DATA

PLASMID	Tm INSERT BEFORE REASSOCIATION	Tm INSERT AFTER REASSOCIATION [a]	Tm [b]	% DIVERGENCE [c]
pAMW1131	83.6	83	-0.6	0.6
1184	90.5	83	-7.0	7
1199	88	79	-9.0	9

a. REASSOCIATION TO COT 500 (SEE TEXT)

b. $Tm = Tm_{AFTER} - Tm_{BEFORE}$

c. % DIVERGENCE : REDUCTION OF Tm BY $1^{o}C$ IS EQUIVALENT TO 1% MISMATCH (BONNER et al, 1973)

low copy number the data might suggest that this is a relatively novel family. If this were so one might expect that the sequences would be clustered in the genome (see Smith, 1975, for example). Alternatively sequence divergence and amplification may be constrained. Aspects of these alternatives can be addressed by hybridisation of the clone to genomic DNA cleaved with restriction enzymes (see later).

The family represented by clone 1184 exhibits approximately 7% divergence by this method of analysis. This is somewhat less than that seen for bulk repeat DNA (10%, see Chapter 3) and may in part explain the paucity of restriction sites generated within the tandem array of repeats. The divergence of the repeats may lead to an underestimate of the copy number (see Appendix II) however this has not been corrected for here. The effect of the %G+C of the clone on rate of reassociation is negligible.

The sequences represented by clone 1199 appear to have a divergence similar to that of bulk repeat DNA. The degree of divergence shown has been demonstrated by others (e.g. Bonner et al, 1973) to cause a twofold reduction in rate of reassociation and hence a lower apparent copy number. However given the heterogeneity in the derivative melting profile it is not clear as yet whether the reduction in Tm can reasonably be attributed to the whole genomic insert. It may be therefore that parts of the cloned sequence are present at a higher copy number than originally deduced.

The restriction mapping, hybridisation and denaturation data all tend to suggest that the three clones represent distinct families within the genome. To demonstrate further that this is so the clones were cross-hybridised against each other by Southern Hybridisation, as follows :

As the inserts cannot be excised by BamHI the whole clone was digested with BamHI and HinfI to separate the vector and insert sequences into readily recognisable fragments. The fragments were size fractionated by agarose gel electrophoresis and transferred to nitrocellulose. Duplicate

filters were probed with a variety of the clones and pAT153. The fragments generated from the vector acted as internal size standards to identify insert fragments, and also acted as a crude indicator of the efficiency of hybridisation and/or transfer of fragments of various sizes. Fig. 5.12 shows the example of clone 1131 hybridised to pAT153, pAMW 1114, 1131, 1184, 1199. The stained agarose gel is shown in **1** while the resulting auto-radiograph is shown in **2.** The hybridisation and washes were done at 60°C in 3 x SSC to try to detect poor homologies between the insert sequences. The arrows indicate positions on the auto-radiograph where cross hybridisation would be expected to be visible if it did occur. There was no noticeable cross hybridisation between any of the clones at the stringency used. This tends to confirm that the cloned sequences are entirely distinct.

Attempts were made to determine whether the clones were internally heterogeneous with respect of repetition frequency in a manner similar to that of Wensink et al, (1978). The experiments (not shown) were inconclusive in that it appeared that the fraction of homologous sequences in the probe used (Cot 100 duplex DNA, S_1 treated and nick translated) was not sufficient to give clear hybrids. In particular when hybridisation to HinfI, HaeIII, HpaII or TaqI digests of the clones was performed it was difficult to detect hybridisation to the smaller fragments, even after long exposures. The problem might be overcome by hybridising sub-fragments of the clones to total genomic, or isolated repetitive DNA in a manner analogous to "gene copy number" experiments.

The organisation of these cloned sequences within the axolotl genome was analysed by Southern hybridisation to restriction endonuclease digested genomic DNA. For this purpose for each enzyme used a large amount of axolotl DNA (50 µg) was mixed with 1 µg 'phage λ DNA and digested to completion. Completeness of digestion was monitored by gel fractionating a small aliquot (equivalent to 1 µg axolotl DNA) of each digest and probing

filters were probed with a variety of the clones and pAT153. The fragments generated from the vector acted as internal size standards to identify insert fragments, and also acted as a crude indicator of the efficiency of hybridisation and/or transfer of fragments of various sizes. Fig. 5.12 shows the example of clone 1131 hybridised to pAT153, pAMW 1114, 1131, 1184, 1199. The stained agarose gel is shown in **1** while the resulting auto-radiograph is shown in **2.** The hybridisation and washes were done at 60°C in 3 x SSC to try to detect poor homologies between the insert sequences. The arrows indicate positions on the auto-radiograph where cross hybridisation would be expected to be visible if it did occur. There was no noticeable cross hybridisation between any of the clones at the stringency used. This tends to confirm that the cloned sequences are entirely distinct.

Attempts were made to determine whether the clones were internally heterogeneous with respect of repetition frequency in a manner similar to that of Wensink et al, (1978). The experiments (not shown) were inconclusive in that it appeared that the fraction of homologous sequences in the probe used (Cot 100 duplex DNA, S_1 treated and nick translated) was not sufficient to give clear hybrids. In particular when hybridisation to HinfI, HaeIII, HpaII or TaqI digests of the clones was performed it was difficult to detect hybridisation to the smaller fragments, even after long exposures. The problem might be overcome by hybridising sub-fragments of the clones to total genomic, or isolated repetitive DNA in a manner analogous to "gene copy number" experiments.

The organisation of these cloned sequences within the axolotl genome was analysed by Southern hybridisation to restriction endonuclease digested genomic DNA. For this purpose for each enzyme used a large amount of axolotl DNA (50 μg) was mixed with 1 μg 'phage λ DNA and digested to completion. Completeness of digestion was monitored by gel fractionating a small aliquot (equivalent to 1 μg axolotl DNA) of each digest and probing

LEGEND TO FIG. 5.12

PLASMID CROSS HYBRIDISATION

Several recombinant clones were digested with the enzymes BamHI and HinfI as described in the text. The fragments were separated on agarose gels and transferred to nitrocellulose (Southern, 1975). Separate identical blots were probed with a single cloned DNA, labelled with ^{32}P by nick translation (Rigby et al, 1977).

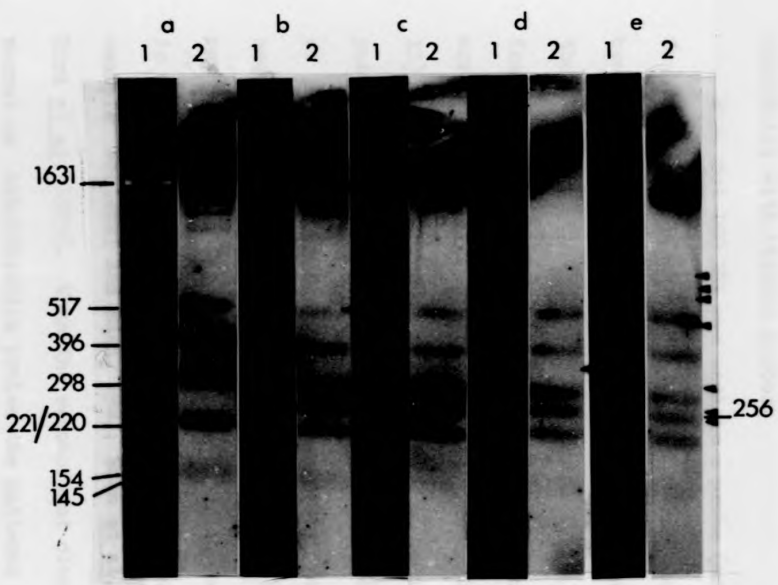
Fig. 5.12 shows the result when pAMW1131 was used as probe.

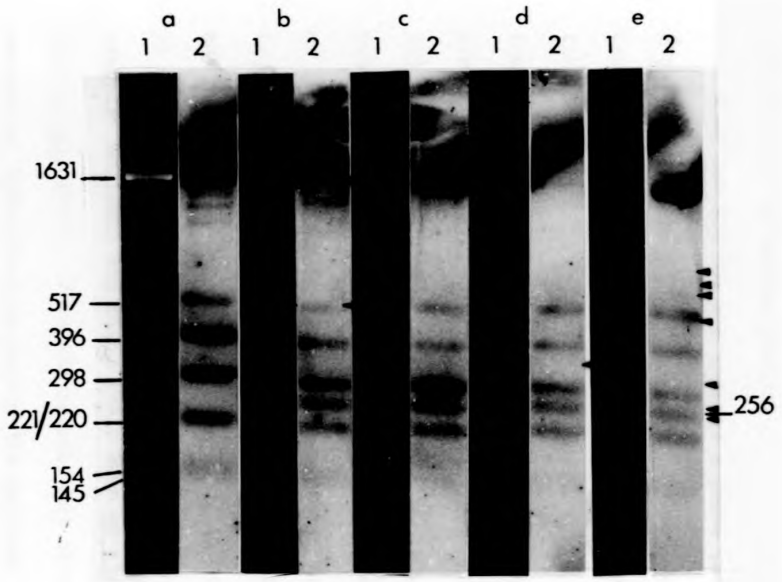
For each of a - e 1, represents the stained gel
2, the resultant autoradiograph

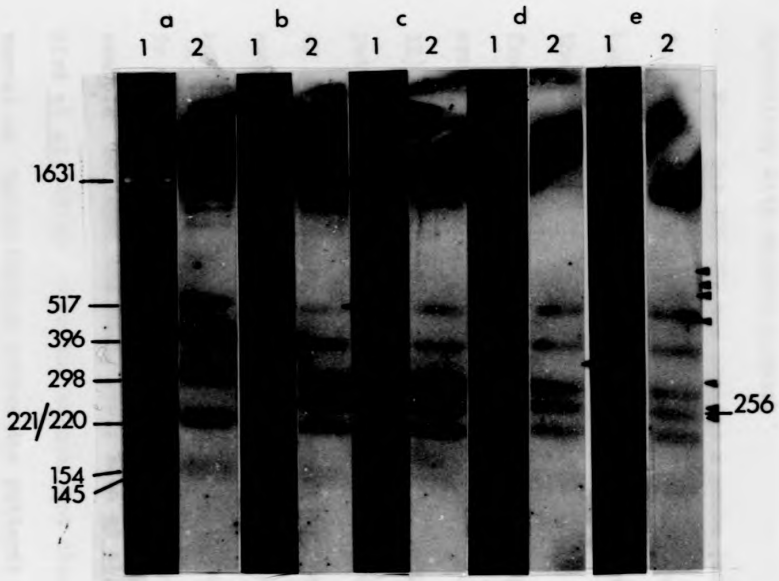
a - pAT153 - HinfI
b pAMW 1114 - HinfI + BamHI
c pAMW 1131 - HinfI + BamHI
d pAMW 1184 - HinfI + BamHI
e pAMW 1199 - HinfI + BamHI

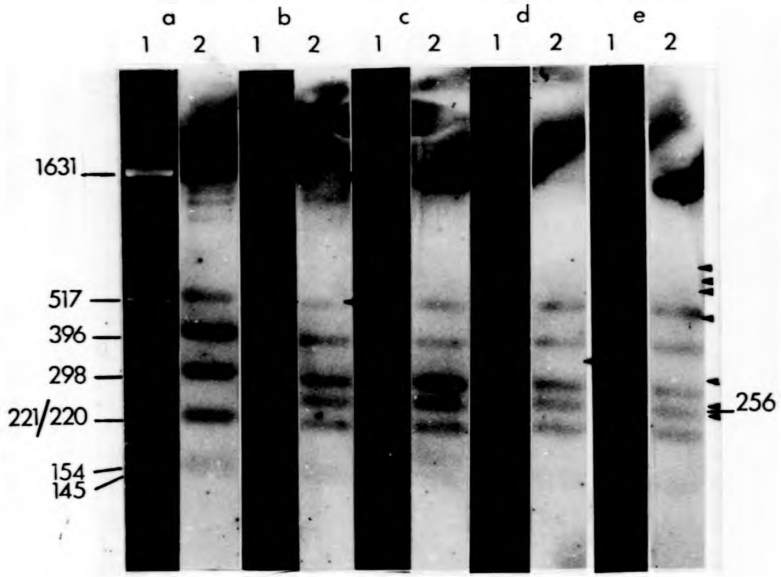
The autoradiograph is a 16 hour exposure, -70°C + intensifying screen.

Fragment sizes for pAT153 - HinfI were calculated from Sutcliffe, 1978 and are bp. The 256 bp band is the Bam^{375} - $HinfI^{631}$ pAT153 fragment.


a
b
c
d
e
1
2
1631
517
396
298
221/220
154
145
256


a
b
c
d
e
1
2
1631
517
396
298
221/220
154
145
256


a
b
c
d
e
1
2
1631
517
396
298
221/220
154
145
256


a
b
c
d
e
1
2
1631
517
396
298
221/220
154
145
256

with ^{32}P λ DNA. This process was particularly important in determining that the Hpa II and MspI digests were equivalent so that methylation at CCGG could be monitored. Additional controls were done to show that the cloned DNA preparations, or the vector pAT153 did not cross-hybridise with the λ DNA, and that the pAT153 did not cross-hybridise with the axolotl genomic DNA.

For the main experiment 5 μg aliquots of each restriction digest were electrophoresed on 1% agarose, transferred to nitrocellulose and hybridised with various probes.

When Cot 100 DNA was used as a probe (not shown) hybridisation to all fragment sizes was seen for all of the enzymes used. No major pattern of bands suggestive of tandem repetition of sequences could be detected above the general level of hybridisation, although the possibility that a minor fraction exists in this organisation cannot be ruled out. The observations are consistent with a wide distribution of Cot 0-100 (Class I; see Section II) sequences throughout the genome. Comparison of the hybridisation patterns to Hpa II and Msp I digests suggests that a substantial fraction of these Class I repeats may be methylated, at least within the 5'-CCGG sequence. Recently several other groups have shown that repeated sequences may be methylated and that the methylation pattern of a particular sequence is maintained after translocation of a sequence to a new location (see for example Sobieski and Eden, 1981; Eden et al, 1981; Whittaker et al, 1981; Bird et al, 1979). When the recombinant clones were used in a similar manner as hybridisation probes the patterns shown in Fig. 5.13 to 5.15 were obtained for clones 1131, 1184, 1199 respectively.

Several features of the genomic organisation of these recombinants can be deduced from the hybrids.

Fig. 5.13 shows the Southern hybridisation of pAMW1131 to genomic restriction digests. It is clear that the axolotl sequences represented in this clone have a complex organisation within the genome. The axolotl DNA

LEGEND TO FIGS. 5.13 - 5.15

GENOMIC HYBRIDISATION OF CLONED DNAs

Axolotl genomic DNA was digested with various restriction enzymes as described in the text. 5 μg of each digest was electrophoresed on 1% agarose gels. The DNA fragments were transferred to nitrocellulose (Southern, 1975).

Separate blots were probed with individual cloned DNAs, labelled with ^{32}P by nick translation (Rigby et al, 1977).

Fig. 5.13 pAMW 1131 hybridised to genomic DNA

5.14 pAMW 1184 hybridised to genomic DNA

5.15 pAMW 1199 hybridised to genomic DNA

A : stained gel B : Autoradiograph

For each gel the lanes represent :

a	Axolotl DNA -	BamHI
b	"	PstI
c	"	EcoRl
d	"	HindIII
e	λ DNA	HindIII
f	Axolotl DNA	HpaII
g	"	MspI
h	"	TaqyI
i	"	HaeIII
j	"	HinfI

Fragment sizes noted are kilobasepairs.

Exposure times : 5.13B 21 days, -70°C no screen; 5.14B 3 days -70°C; Fig. 5.15B 6 days -70°C.

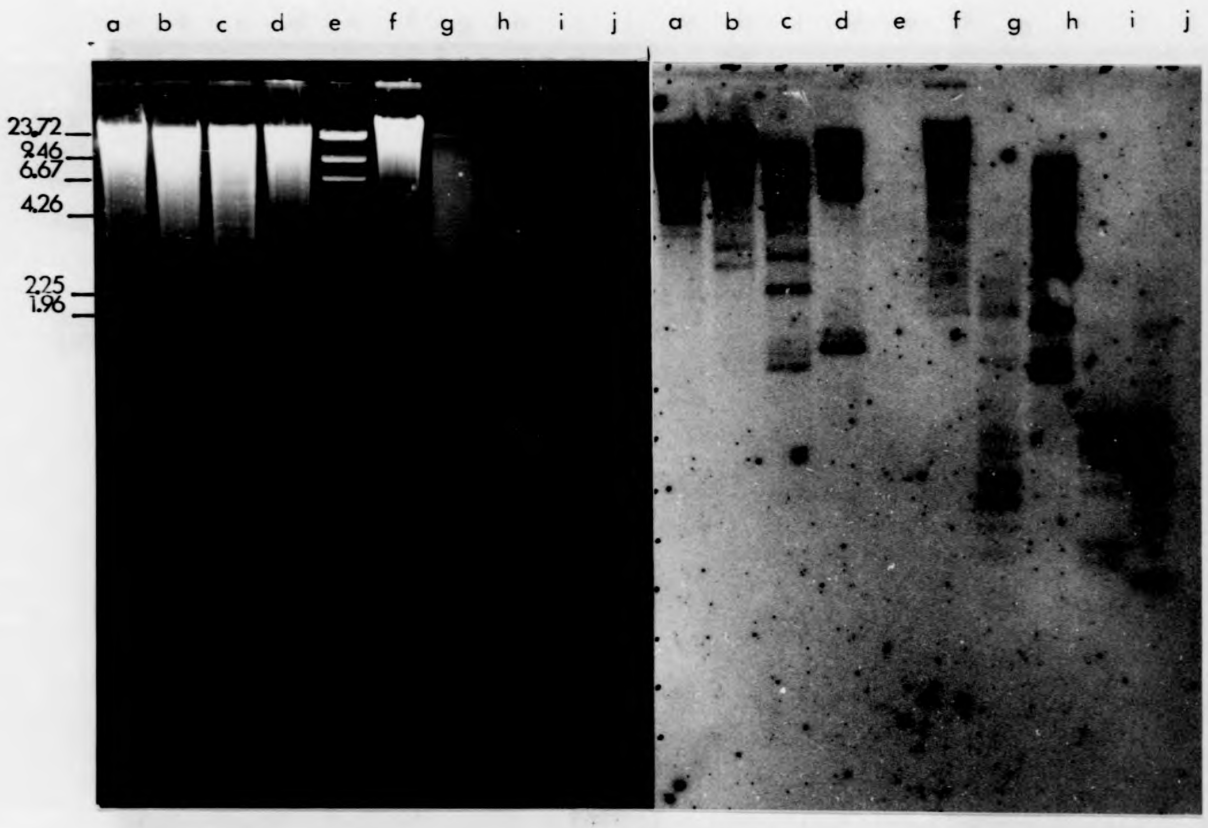
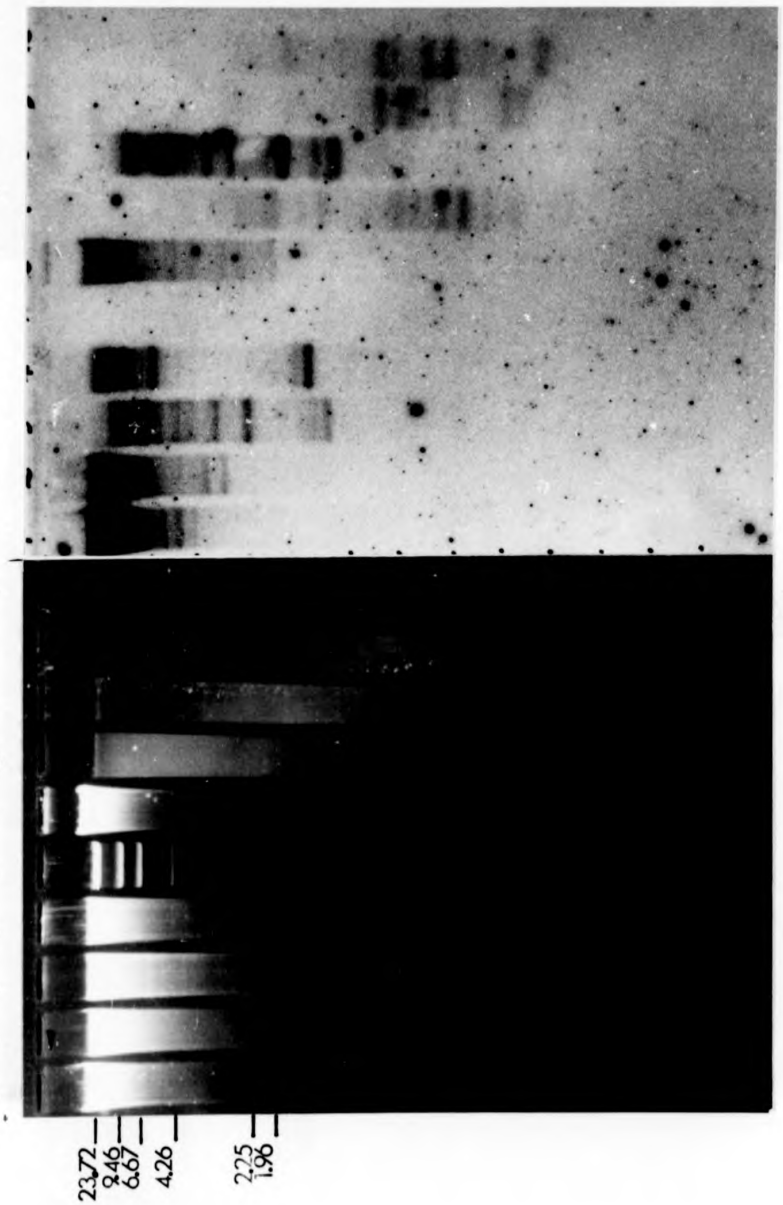


Fig 5.13

Fig 5.13



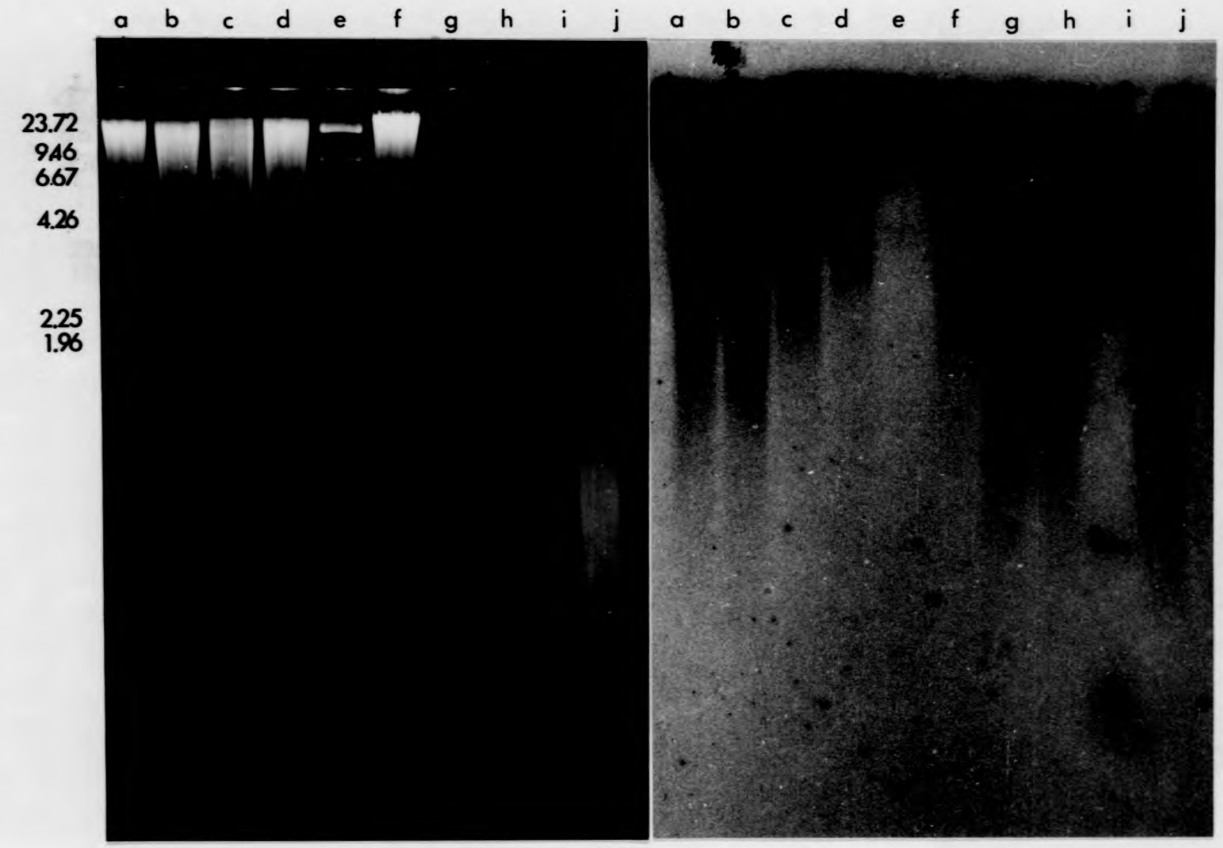


Fig 5.14

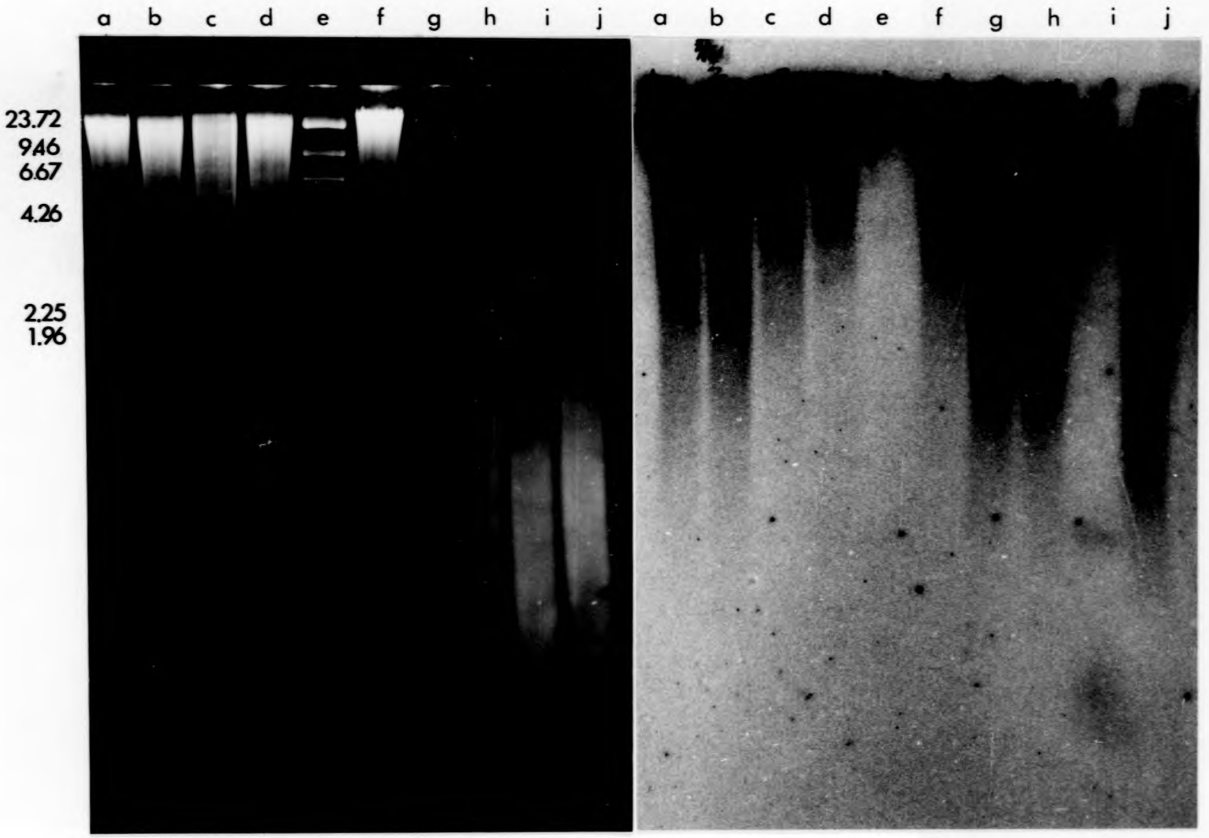


Fig 5.14

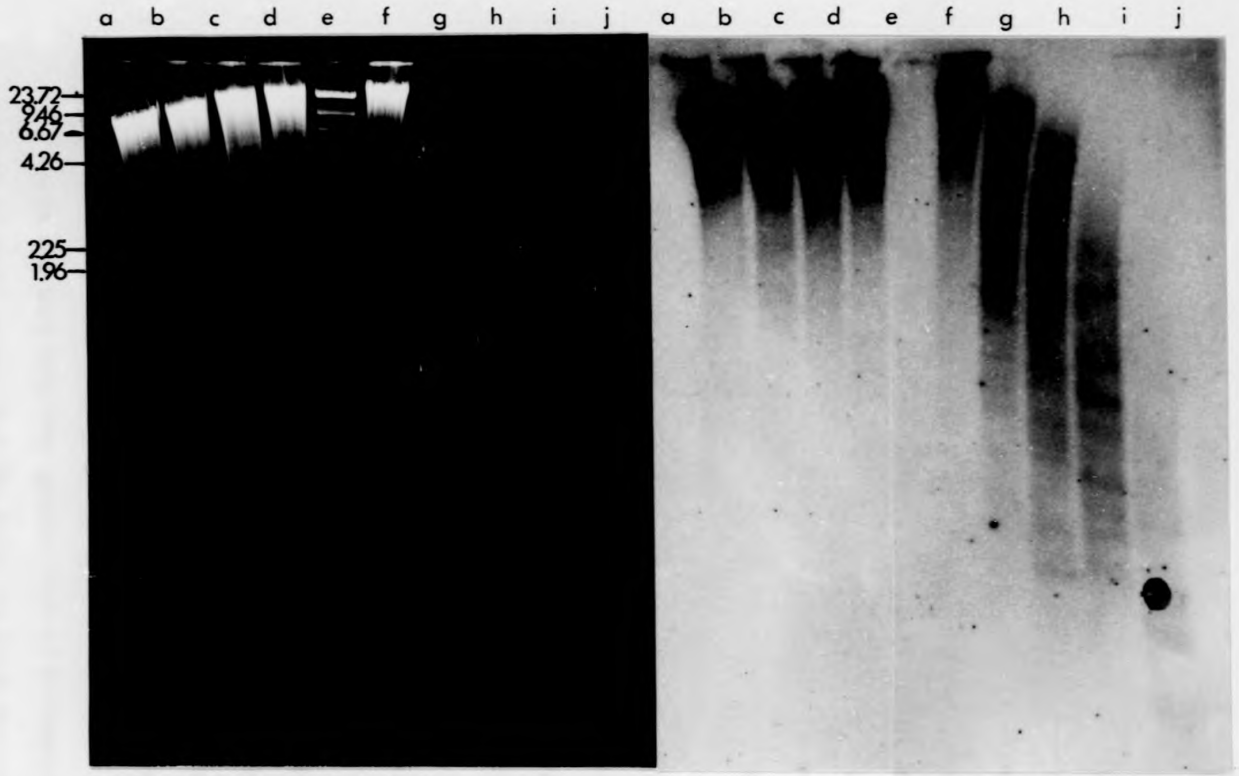


Fig 5.15

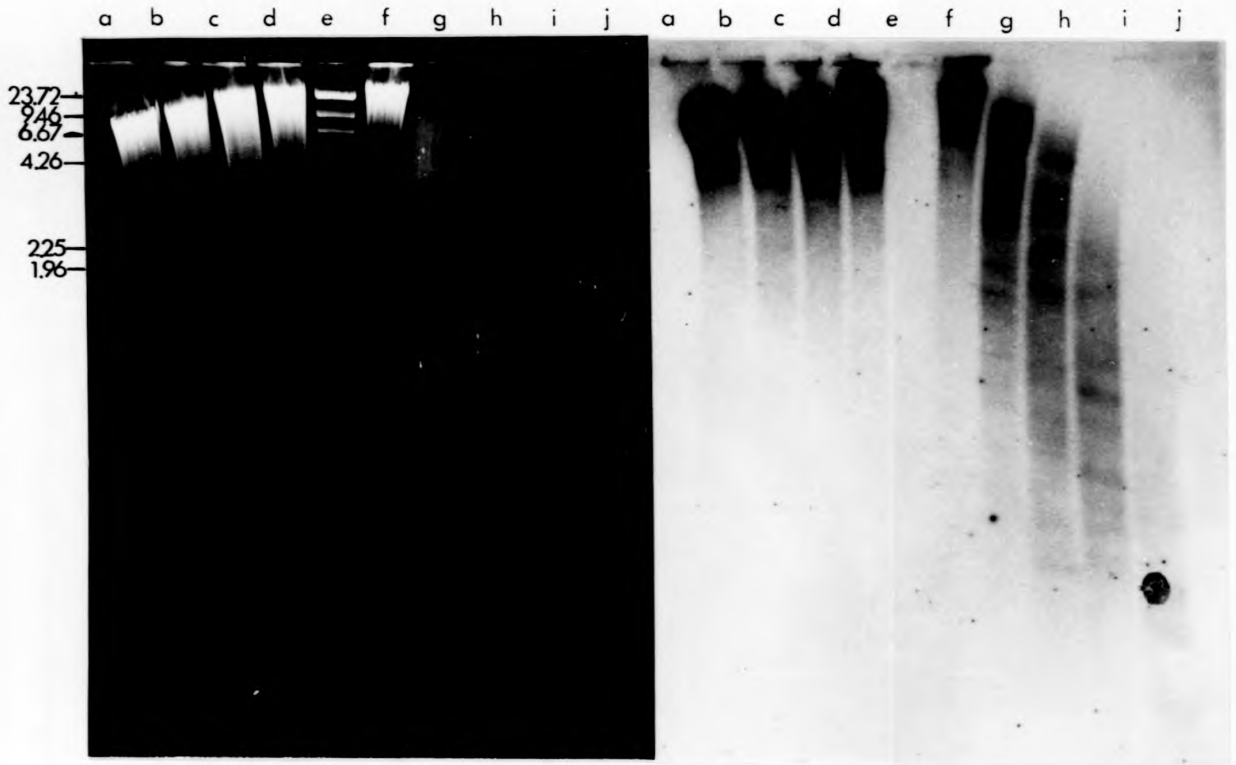


Fig 5.15

used in this particular experiment was obtained from the blood of a single animal so that the multiple bands are not simply due to the effects of having DNA from different individuals or different tissues of the same animal. It is not possible to derive a series of bands based on multiples of some unit length for any of the restriction enzymes. The different intensities of the bands at different fragment sizes suggest that the sequence represented by the clone has undergone reduplications.

When the enzymes recognising tetranucleotide sites are considered it can be seen (with the exception of Hpa II) that there appears to be little background smearing. This suggests that the short range organisation of the sequence in the genome is fairly constant. Thus the axolotl sequence in pAMW1131 may be present in a limited number of sequence backgrounds which have undergone subsequence reamplification as general interspersion of a small sequence such as 1131 would inevitably lead to heterogeneity in the location of adjacent restriction sites. These putative larger sequences cannot be too large however as the enzymes which recognise hexanucleotide cutting sites tend to produce a smear suggesting heterogeneity of adjacent sequences. This also suggests that the sequence is dispersed throughout the genome.

A further feature of interest is the contrast between the Hpa II (f) and Msp I (g) hybridisation patterns. It would appear that the majority of the sequences which hybridise to clone 1131 are methylated within the 5'-CCGG recognition tetranucleotide, as they are resistant to Hpa II and sensitive to Msp I (see also Section I). Some bands are however present in both tracks which suggests that the probe sequence is bordered in these regions by non-methylated 5'-CCGG sites.

Taken together with the apparent lack of divergence between family members it would appear that the family to which the insert in 1131 belongs is either new in evolutionary terms or the sequence has been conserved, whether by selection for some function or by a form of gene conversion

(see for example Brown and Dover, 1981). Comparative studies between species should provide a means of discriminating between these alternatives.

Fig. 5.14 shows the Southern hybridisation of clone pAMW1184 to restriction fragments of axolotl DNA. For the most part the probe hybridises to sequences across the whole range of fragment length. This could be indicative of a general interspersion throughout the genome of the sequence represented by the cloned insert. However the HaeIII (i) pattern is noticeably different. Here the hybridisation is confined to several bands. These are not clearly resolved on 1% agarose gels, however when 2.5% agarose gels were used to separate the restriction fragments the bands were more clearly visible and corresponded to the internal HaeIII fragments found when the clone is digested with HaeIII (see Fig. 5.6A). It would thus appear that there is a regularly repeating internal organisation. Therefore the heterogeneity of the hybridisation when other enzymes are used to digest the axolotl DNA may simply be due to the infrequent location of restriction sites within a tandem array. To use the terminology adopted for satellite DNA the HaeIII pattern would be an 'A' pattern while the rest of the enzymes used would show a 'B' pattern (Horn and Zachau, 1977). Given the regularity of HaeIII sites, both within the cloned sequence and in the genome, and the highly repetitive nature of the clone the possibility arises that 1184 is a representative of a satellite DNA fraction. Such a fraction was not identified in earlier experiments (see Section I). However, even supposing the 57 bp repeat represented % of the genome (Table 5.3) it is highly likely that such a fraction would be undetected in the analyses performed so far, although the family may be a member of the % G+C rich shoulder seen when 2.5 kb DNA is subjected to analytical CsCl ultracentrifugation (Fig. 3.7). There is a formal possibility that multiples of the monomer sequence may be interspersed amongst other sequences within the genome. This would explain the heterogeneity of hybridisation seen with most restriction enzyme digests. Recently

Spohr et al (1981) have isolated a short repeat of 80 bp in Xenopus laevis. This repeat has 10^5 copies per haploid genome and is organised as small groups of the basic monomer interspersed throughout the genome. In Xenopus this repeat is transcribed in both oocytes and stage 40 embryos. It is possible that the sequence represented by clone 1184 may be organised in a similar manner.

When the Hpa II and Msp I hybridisation patterns are examined it can be seen that the majority of labelling in the Hpa II track is of higher molecular weight than that in the Msp I track. This suggests that the regions surrounding the 1184 sequences and probably the 1184 sequences themselves are methylated in blood cell nuclei. The implications of this observation will be discussed later with respect to the transcriptional activity of the cloned sequences.

Fig. 5.15 shows the Southern hybridisation of clone pAMW1199 to restriction fragments of axolotl DNA separated by size on agarose gels. It is immediately obvious that, above the general level of hybridisation, there are a number of bands in the lanes corresponding to digests by enzymes recognising tetranucleotide sites (g - l). These bands suggest that there is some constancy in the organisation of the 4.9 kb fragment throughout the genome. However there is also a background smearing which may indicate the heterogeneity of flanking sequences. There is the possibility that the smearing also represents some internal rearrangements which act to vary the distance between particular internal cutting sites (or remove them altogether) in individual sequences (see for example Musti et al, 1981).

The hybridisation to digests by enzymes recognising hexanucleotide sites (a - d) indicate that the sequences represented in clone 1199 are present in a variety of different sequence surroundings throughout the genome.

As has already been shown for the other clones already described the Hpa II/Msp I patterns shown by 1199 indicate extensive methylation of sequences in and around the cloned sequence in the genome. It would appear,

based on the observations with the cloned axolotl DNA (above), bulk DNA (Section I) and total Class I repeats (not shown) that axolotl DNA is extensively methylated. This is in agreement with the findings of Bird (1980) on amphibian DNA methylation levels.

The tissue used to prepare DNA for these experiments was red blood cells. It has been suggested that during differentiation DNA sequences become demethylated from a level seen in the germ cells. The demethylation may potentiate or be a prerequisite of gene expression. However, it has also been suggested that the required demethylation need only involve a subset of methylated sites in and around control regions. Thus for example X. laevis ribosomal DNA can be expressed even while the majority of methylated sites are still present (Bird et al, 1981).

Whether bulk repeat DNA or repeat sequences represented by the clones described here were capable of being transcribed in vivo despite the degree of methylation observed, was investigated below. It has been known for some time that a fraction of the repetitive portion of the genome of several species is transcribed. While some of this transcription will be due to repeated genes, e.g. histones, rRNA, 5SRNA, the majority cannot be ascribed to these genes. The presence of repetitive DNA in hnRNA has led to suggestions of a role in gene regulation for these sequences (e.g. Davidson and Britten, 1979), although the observation that the repetitive DNA may be located in introns (e.g. Ryffel et al, 1981; Heilig et al, 1980) needs to be taken into consideration.

Recently repetitive sequences have been located adjacent to coding sequences on mRNA or cDNA made against poly A^+RNA in human (Crampton et al, 1981), sea urchin (Constantini et al, 1980) and Dictyostelium (Kimmel and Firtel, 1979). In Dictyostelium the transcripts appear to be asymmetric, however in the sea urchin it has been shown that both strands are transcribed although the evidence suggests that one strand is preferentially transcribed (Moore et al, 1978, 1980). Very recently it has been shown that highly

repetitive DNA, located between histone clusters in the newt Notopthalmus, is transcribed, at least in the oocyte, and that transcription occurs from both strands of the repetitive sequences, again in different amounts (Diaz et al, 1981; Stephenson et al, 1981). The amphibian oocyte has provided much evidence for the transcription of repeated sequences, from the early kinetic studies on oocyte RNA complexity in Xenopus (e.g. Hough and Davidson, 1972) to the many in situ hybridisation studies using repeated DNA to hybridise to nascent RNA on the loops of lampbrush chromosomes (e.g. Barsacchi and Gall, 1972; MacGregor and Andrews, 1977; MacGregor, 1979; Varley et al, 1980 a + b; Diaz et al, 1981; Vlad and Hilder, in press).

However several points need to be noted. First, although lampbrush chromosomes have been found to be actively synthesising RNA and have been assumed to be the site of synthesis of maternal RNA (reviewed in Davidson, 1976; Sommerville, 1977; MacGregor, 1980) it has been estimated that only 5% of the genome is transcribed during the lampbrush stage (Callan, 1963).

Second, the observations of "read through transcripts" that is transcription past putative transcription termination signals, to give very long transcripts (Sommerville and Malcolm, 1976; Varley et al, 1980 a + b; Diaz et al, 1981) and transcription using "incorrect polymerases" (Morgan et al, 1980) suggests that transcription during the lampbrush stage may not be faithful. In this respect Golden et al (1980) have observed that poly A^+ RNA accumulates early in oogenesis and then maintains a steady state throughout the lampbrush stage. Individual RNAs probed using cDNA clones of oocyte poly A^+RNA do not appear to change in level throughout the lampbrush stage even after the maximal level of accumulation has occurred. Ford et al (1977) investigated the stability of early poly A^+RNA and suggest that such RNA is stable throughout oogenesis. Golden et al suggest that lampbrush chromosomes transcripts may be metabolically unstable and not equivalent to the population of poly A^+RNA synthesised in pre-vitellogenic oocytes. Such RNA may have a simple space filling function

designed to swell the germinal vesicles (Cavalier-Smith, 1978).

However other evidence suggests that poly A^+RNA is made during the lampbrush stage (Dolecki and Smith, 1979) and that lampbrush transcripts are actively transported into the cytoplasm and have the same sequence composition as oocyte cytoplasmic RNA (maternal RNA) (Anderson et al, 1982). These authors suggest that oocyte cytoplasmic RNA turns over fairly slowly yet the RNA pool is so large that lampbrush transcription is required at maximal rates to maintain a steady state of cytoplasmic RNA through oogenesis.

It is evident that the function and consequences of lampbrush transcription are not yet clear, however each of the hypotheses outlined above, if proved, would provide a clue as to the function of repeated sequences in the genome. It was therefore of interest to determine whether the Class I repeats and the cloned DNAs described above were transcribed on lampbrush chromosomes.

In order to demonstrate that axolotl lampbrush preparations could be analysed in this way control experiments using a plasmid containing the X. laevis 18 and 28 S rDNA and adjacent spacers (pXl 101, provided by J. Coveney) were performed. In these controls hybridisation could be seen to the amplified nucleoli found in oocyte preparations. However in all the preparations studied hybridisation could not be detected at the region of the proposed nucleolus organiser on chromosome III (Callan, 1967). A similar situation has been demonstrated in Triturus oocytes (Morgan et al, 1980), in which extra chromosomal ribosomal RNA genes are actively transcribed, but those located within the nucleolus organisers on the chromosome are not. It would however appear that there is nothing in the oocyte preparations which may prevent hybridisation using axolotl DNA.

The Class I repeat probe was prepared by reassociating sheared DNA to ECot100 followed by S_1 nuclease digestion to remove single strand tails, a proportion of which would be adjacent slower reassociating sequences.

The DNA was nick-translated using all four ^{3}H deoxynucleotide triphosphates and hybridised to lampbrush preparations.

Fig. 5.16 shows various preparations hybridised with the Cot probe. In most cases the grain density over the loops and axes was higher than the surroundings. This might suggest a diffuse hybridisation over the whole chromosome set. However an alternative hypothesis is that the probe is merely sticking to chromosomal proteins. This possibility was not tested. In some instances specific, intense, hybridisation can be seen on particular loops (see for example 5.16a). In other cases hybridisation begins at a point along the loop and increases in grain density (Fig. 5.16b). In several instances hybridisation occurs on apparently small loops over the axis (Figs. 5.16 c - e). Whether the hybridisation is to loops and/or at axial structures e.g. spheres or granules, cannot easily be determined. Considering the fraction of the genome occupied by the Class I repeats (see Chapter 3) there were surprisingly few examples of discrete hybrids of large size over the chromosome sets examined. Fig. 5.16f shows an example of a complete bivalent . It can be seen that there are few major hybrids on this chromosome. The relative lack of hybrid formation may be a consequence of the organisation of the Class I repeats, described earlier, in that most repeats are interspersed amongst other repeats and might not occur within a transcription unit, even on the size scale of a lampbrush loop. Alternatively the observation that much of the Class I repeat DNA, and the cloned DNAs analysed, appear to be methylated may preclude transcription. Another possibility is that as the probe DNA is heterogeneous the concentration of any given sequence in the probe might be too low to obtain a noticeable hybrid in a reasonable length of time. These possibilities are not mutually exclusive.

It is clear that at least some Class I repeats are transcribed during the lampbrush stage of oogenesis. The possibility that the cloned DNAs analysed above (and others not described above) form a subset of these

transcribed repeats was investigated.

Therefore each of the clones described in detail above, and several others noted in Table 5.1 were ^{3}H labelled by nick translation and hybridised as described. In all cases, with the exception of pAMW1184, no convincing hybridisation could be detected, even after extremely long exposures (more than six months in some cases). For pAMW1184 there did appear to be good evidence for hybridisation. However, it must be stressed that in these cases long exposures were required. This may suggest that isolated repeats of the satellite-like 1184 sequence were being identified, rather than long sequence tracts as in the case of, for example, satellite I in Notopthalmus (Diaz et al, 1981). Several examples of the observed hybridisation are shown as Fig. 5.17. Fig. 5.17a shows a section of chromosome VI from preparation M2. The spheres (Callan, 1966) and centromere are easily identified as is a hybrid on a small loop pair on the long arm and a hybrid apparently at a chiasma point. Two further hybrids, again one at a chiasma are located on the shorter arm of chromosome VI (Fig. 5.17b). A number of hybrids appear at or close to the telomeres (e.g. Fig. 5.17c, at the end of chromosome VIII, and Fig. 5.17d, at the end of chromosome II of preparation L18). Karyotype analysis was performed on three preparations by my supervisor Dr. Vlad. The analysis is shown on Fig. 5.18. Several of the hybrids persist even on preparations in which the loops are retracting, of which L18 is an example. Preparation M10 was peculiar in that the major sphere loci are usually seen on chromosomes VI and XIII (Callan, 1966). On M10 the spheres appeared to be on chromosomes II and XIII, the region around the centromere of chromosome II (M10) and VI (M2) appear quite similar. It is not known at the moment whether the hybrids do in fact represent the transcription of long arrays of the 1184 repeat, or merely isolated transposed repeats present in other sequence arrays.

Disappointing results were obtained when hybridisation to axolotl metaphase chromosomes was attempted. Livesey (1980) has described methods

LEGEND TO FIG. 5.16

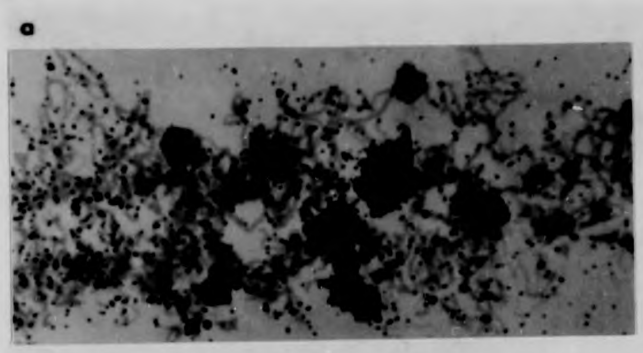
HYBRIDISATION OF CLASS I REPEATS TO AXOLOTL LAMPBRUSH CHROMOSOMES

a - f show examples of various hybrids

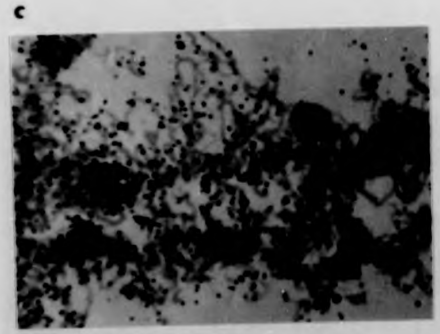
a	Preparation	K10
b	"	N22
c	"	K10
d	"	K12
e	"	K5
f	"	K11

The scale bar represents 50 μm for 5.16 a - e and 10 μm for 5.16f.

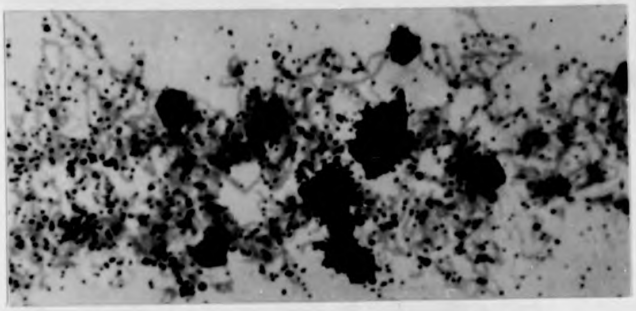
100 ng (4×10^5 c.p.m.) of probe was applied to each preparation in 20 μl of 4 x SSC/50% Formamide and incubated at 37°C for 16 hours. Washing was exactly as in the Methods section.

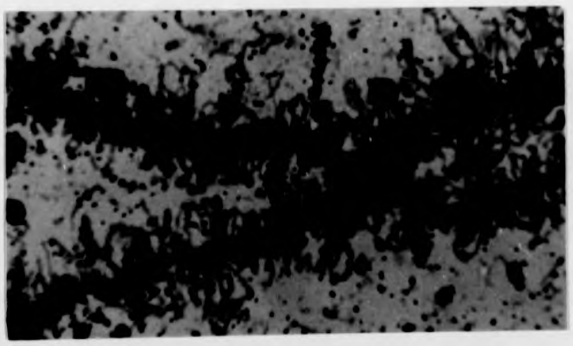

a

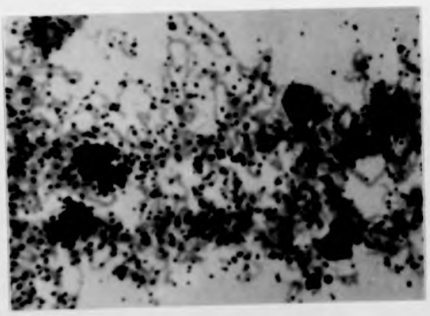

b


c


a
b
c



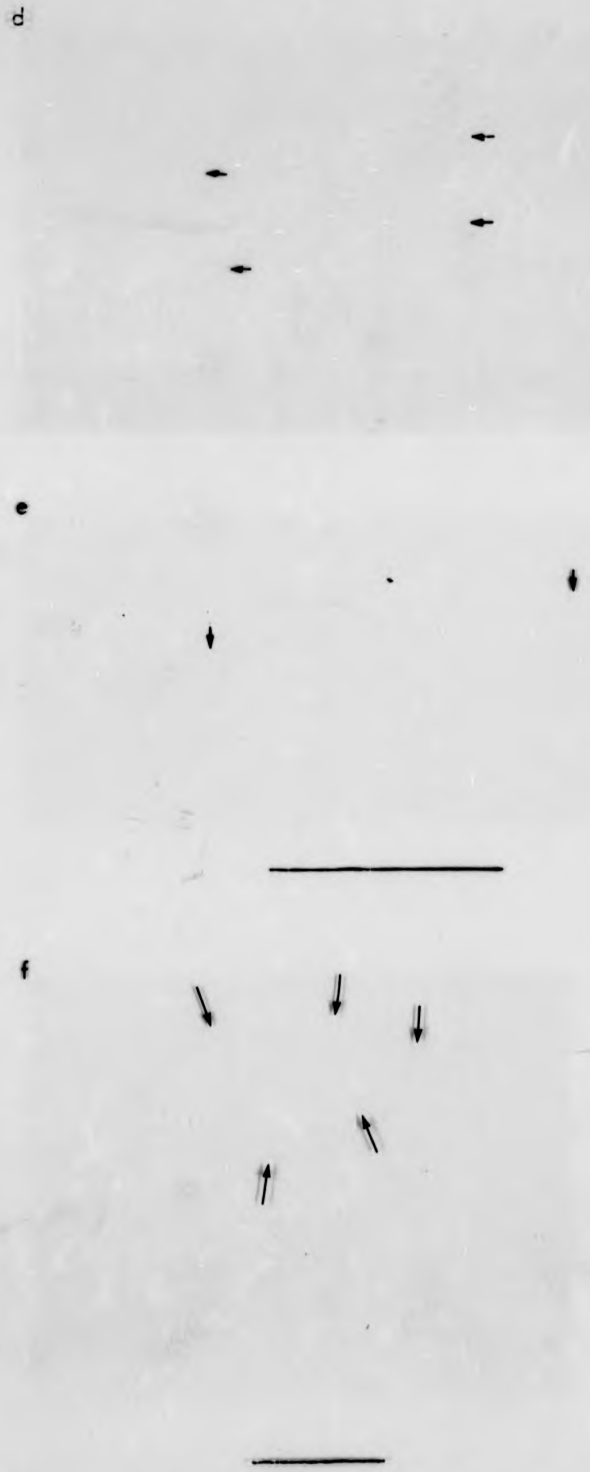


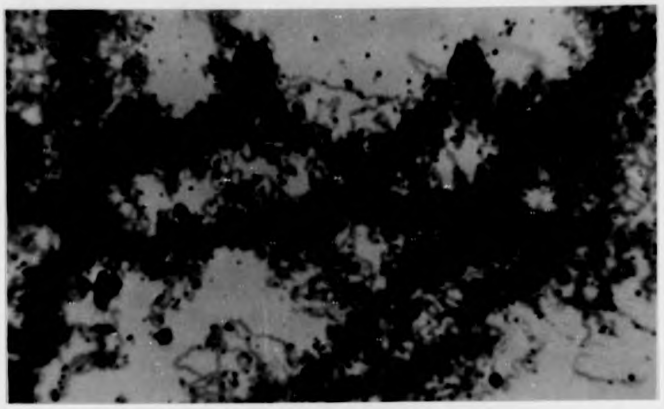


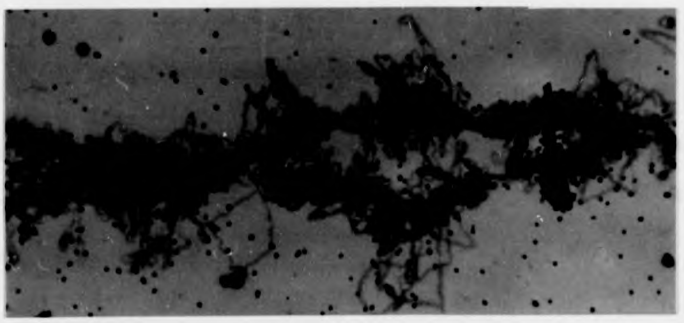
d

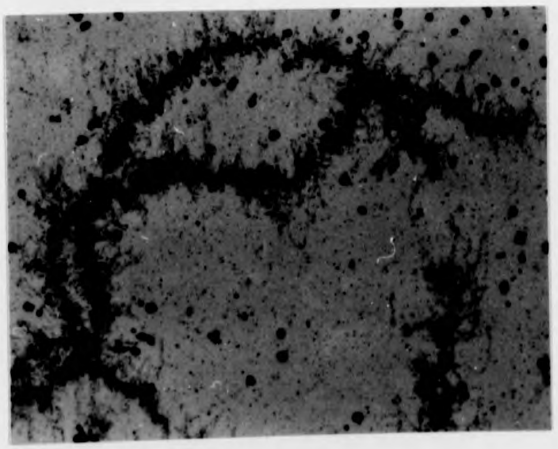
e

f


d
e
f







LEGEND TO FIG. 5.17

HYBRIDISATION OF pAMW1184 TO AXOLOTL

LAMPBRUSH CHROMOSOMES

a - d show examples of various hybrids

a	M2	Chromosome VI
b	M2	Chromosome VI
c	M2	Chromosome VIII
d	L18	Chromosome II

The scale bar represents 50 µm in a, b and c, and 20 µm in d

c = centomeres, s = spheres

100 ng (2×10^5 cpm) of probe was applied to each preparation in 20 µl of 4 x SSC/50% Formamide and incubated at 37°C for 16 hours. Washing was exactly as in the Methods section.

LEGEND TO FIG. 5.17

HYBRIDISATION OF pAMW1184 TO AXOLOTL

LAMPBRUSH CHROMOSOMES

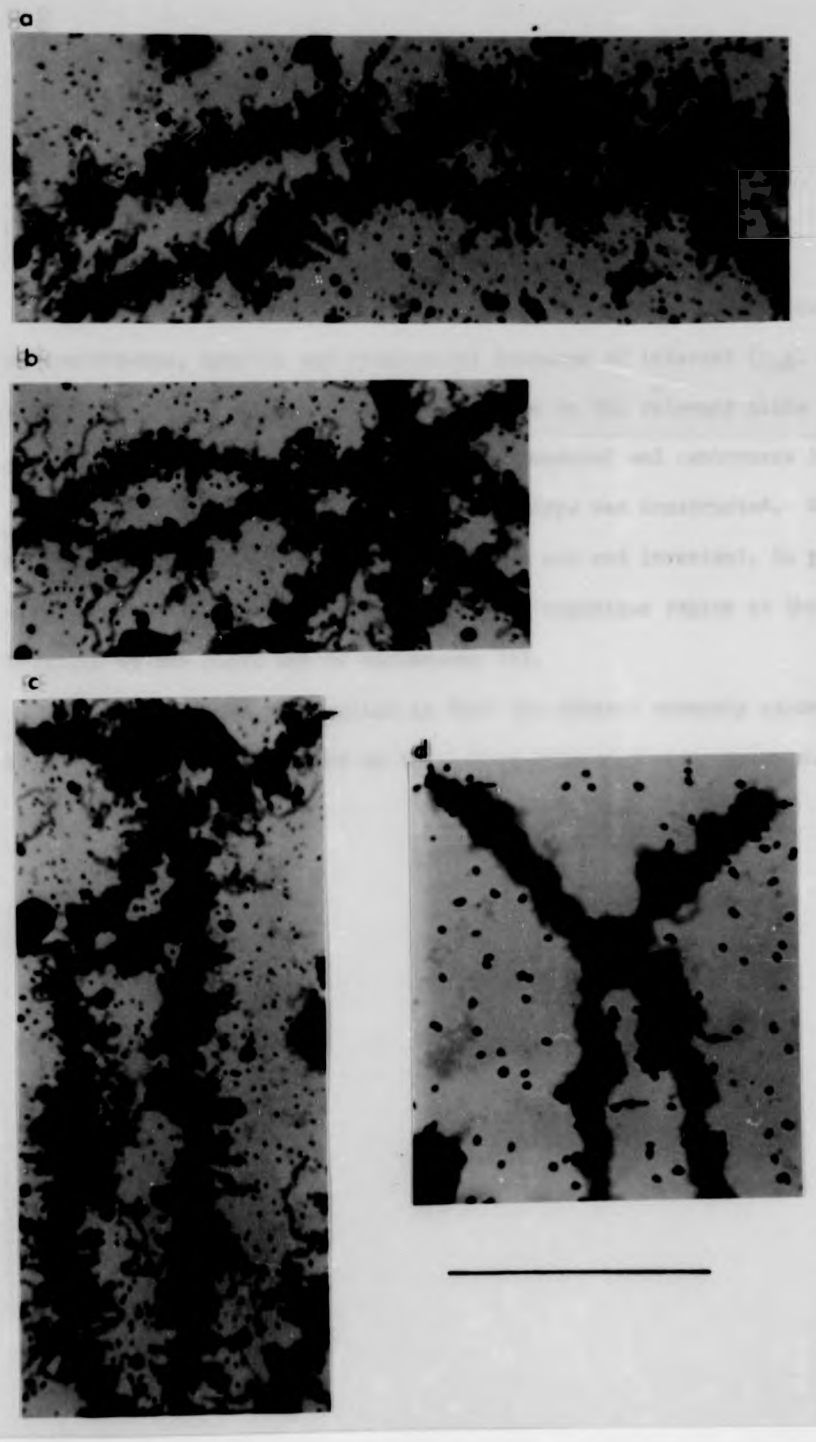
a - d show examples of various hybrids

a	M2	Chromosome VI
b	M2	Chromosome VI
c	M2	Chromosome VIII
d	L18	Chromosome II

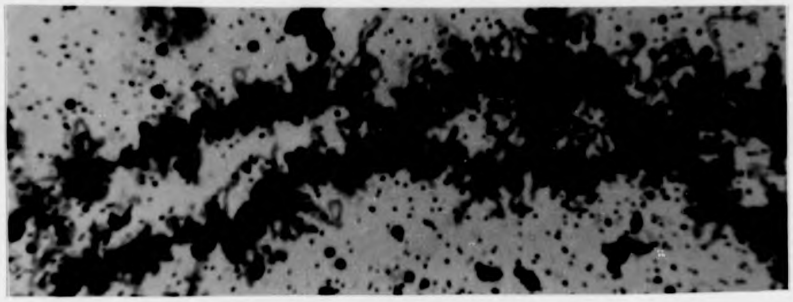
The scale bar represents 50 μm in a, b and c, and 20 μm in d

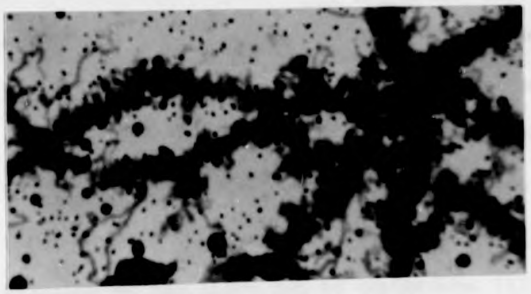
c = centomeres, s = spheres

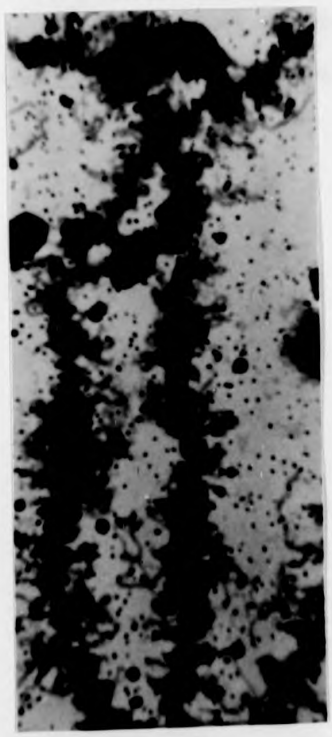
100 ng (2×10^5 cpm) of probe was applied to each preparation in 20 μl of 4 x SSC/50% Formamide and incubated at 37°C for 16 hours. Washing was exactly as in the Methods section.

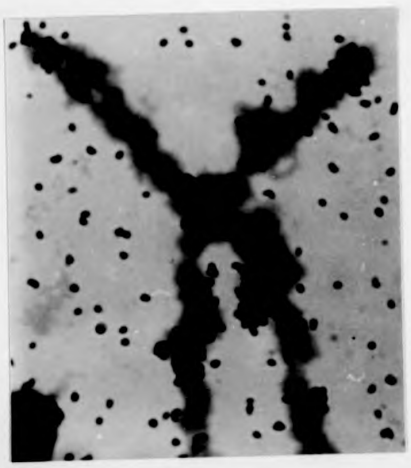

a
b
c
d

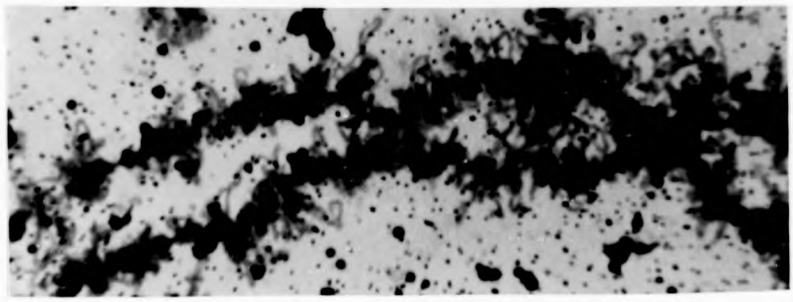

a
s
c
s
c
b
c
d

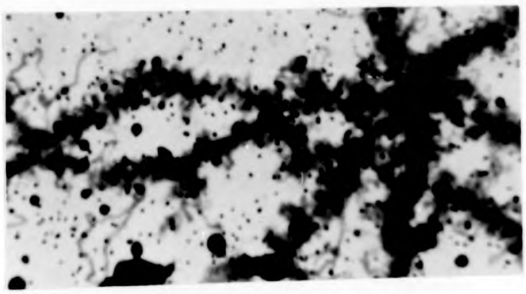


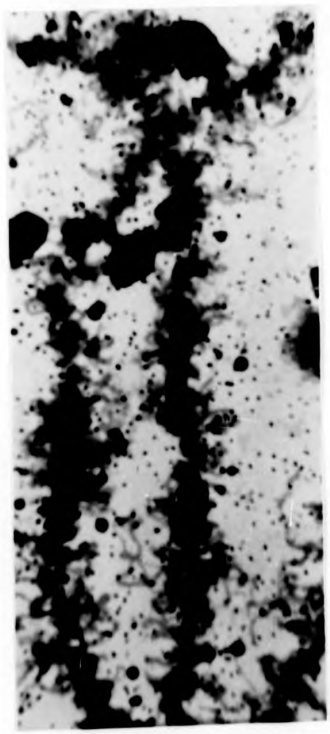












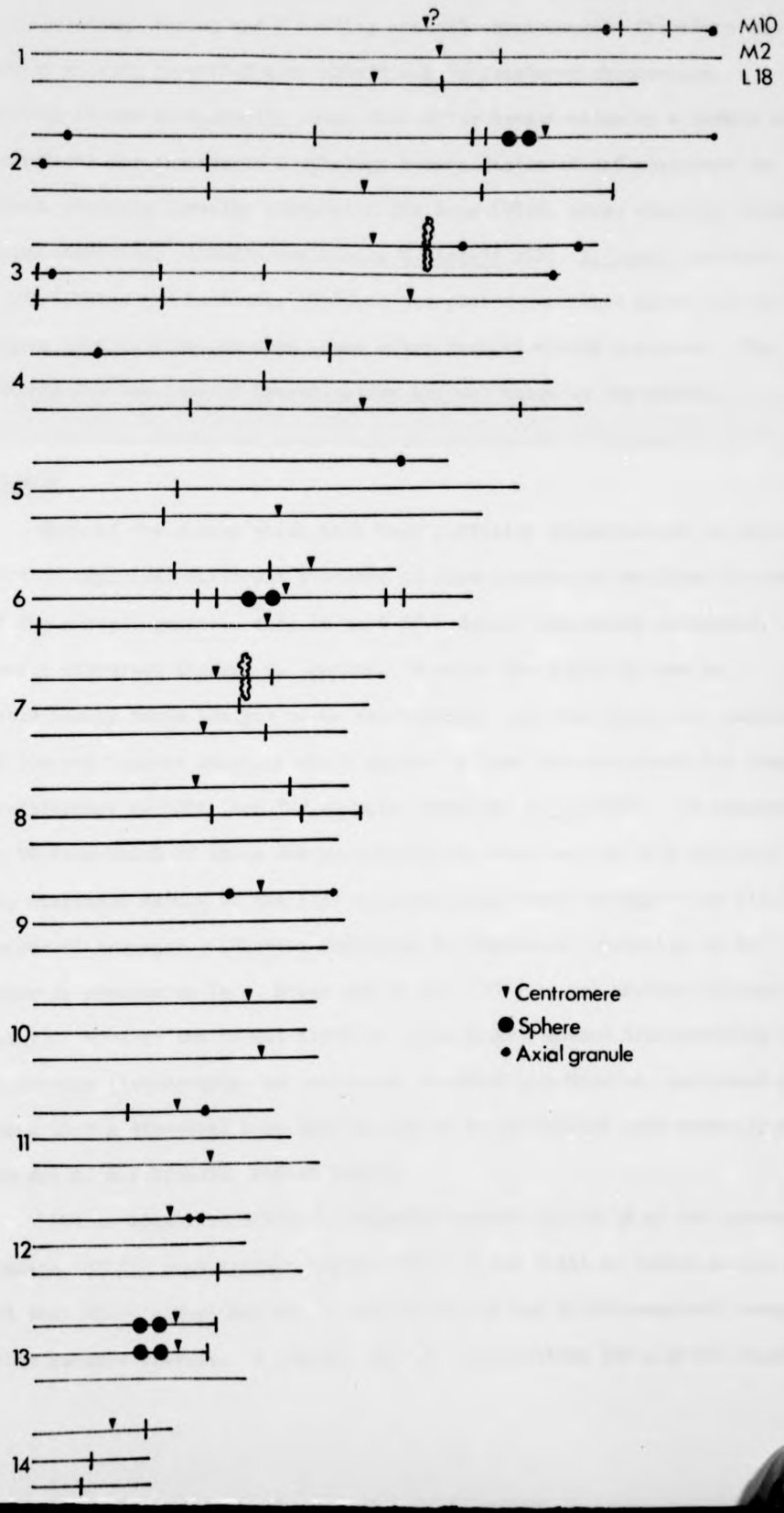
LEGEND TO FIG. 5.18

PARTIAL KARYOTYPES OF LAMPBRUSH CHROMOSOME PREPARATIONS HYBRIDISED WITH ^{3}H pAMW1184

(prepared and included by courtesy of Dr. M. T. Vlad)

Preparations were photographed at X30 magnification and the location of centromeres, hybrids and cytological features of interest (e.g. spheres, axial granules) were positioned by reference to the relevant slide preparation. The chromosome lengths were measured and centromere indices calculated. Based on Callan (1966) a karyotype was constructed. We have noted that the features described by Callan are not invariant, in particular we have not observed a noticeable nucleolus organiser region at the terminus of the short arm of chromosome III.

Preparation M10 is peculiar in that the spheres commonly associated with chromosome VI are found on the second largest chromosome pair.


▼?
M10
M2
L18
1
2
3
4
5
6
7
8
9
10
11
12
13
14
▼ Centromere
● Sphere
• Axial granule

for isolating, fixing and C banding axolotl chromosomes. Therefore the probes already described were hybridised to denatured chromosomes. However it was consistently noted that after denaturation by a number of procedures the chromosome morphology became distorted and resistant to Giensa staining (usually signifying DNA loss (Vlad, pers. comm.)). Under these conditions plasmids containing X. laevis rDNA, X. laevis histone H_4DNA (Turner and Woodland, 1982) or the probes described above did not appear to hybridise, in some cases after several months exposure. The reasons for the lack of hybridisation are not known at the moment.

SUMMARY

Each of the clones which have been partially characterised in this section represent different features of organisation of the Class I repeats of the axolotl genome. 1131 is part of a small, apparently conserved, family dispersed through the genome. Whether the family is new in evolutionary terms has yet to be established. However there are examples of low copy number families which appear to have been conserved for long evolutionary periods (see for example, Scheller et al.1981). It remains to be seen which of these two possibilities describes the 1131 sequence. The dispersed nature of the 1131 sequence might tend to suggest an old, conserved sequence. (Whether conserved by functional necessity or by sequence conversion (e.g. Brown and Dover, 1982) is yet another intriguing point). However the recent plethora of data on sequence transposition in eukaryotes (transposons, retroviruses, orphons, pseudogenes, processed genes) means that a dispersed organisation cannot be correlated automatically with the age of a particular repeat family.

1184 is highly repetitive occupying around 0.4 - 0.5% of the genome. Features of its organisation suggest that it may exist as tandem arrays, but that these arrays may not be extensive and may be interspersed amongst other genomic classes. A similar type of organisation for a short repeat

has been noted in X. laevis (Spohr et al, 1981). The possibility that 1184 is a representative of a true satellite DNA cannot as yet be eliminated. The large size of the axolotl genome means that detection of satellites is difficult. The fragmentary evidence from lampbrush chromosome in situ hybridisations might suggest that some family members are interspersed in the genome and appear on lampbrush transcripts. The transcription of short, highly repetitive sequences has been demonstrated in a number of instances (e.g. Callan and Old, 1980; Varley et al, 1980 a + b; Diaz et al, 1981; Spohr et al, 1981).

Genomic sequences which hybridise to clone 1199 represent around 0.2% of the genome, although this may be a lower estimate as there is some evidence to suggest that diverged sequences exist which could not cross-hybridise at the criterion used. The complex genomic organisation of the sequences represented in clone 1199 suggests that there may be some degree of sub-organisation within the cloned sequence (see for example Musti et al, 1981; Scheller et al, 1981). Thus the 4.9 kb. element may be composed of a variety of sub-elements each with a distinct copy number and degree of divergence, and hence be an example of the interspersed Class I repeats inferred to exist from the data in Section II of these results.

It is clear from these preliminary results that the Class I repeats to which these clones belong are spread through the genome, as suggested from the kinetic data. It should now be possible to use these clones to examine longer range relationships in the axolotl genome, by preparing phage λ or cosmid libraries of large axolotl DNA fragments and using the clones as sequence probes. In this way it should be possible to examine the detailed organisation of these distinct Class I sequences and so develop a picture of the fine organisation of the axolotl genome (e.g. Moore et al, 1981). It should also be possible to use the clones to probe the genomes of the related Ambystomatids (see Appendix I), to infer evolutionary relationships and trends and to establish the rate at which

sequences expand, diverge and are lost in closely related species. (See for example Mizuno and MacGregor, 1974; Moore et al, 1978, 1981; Scheller et al, 1981). This data will be of great value in determining the factors involved in the development of large genomes.

CHAPTER 6

CONCLUDING REMARKS

The work described in this thesis has been directed towards a detailed understanding of the genome of a large genome animal, the axolotl, in particular the amount and organisation of the repetitive fraction of the genome and of individual members of that repetitive fraction.

The axolotl genome has been shown to be qualitatively similar in composition to other high C value salamanders. The analysis has been developed further and the organisation of the major repeat class has been described in some detail.

Fragments of axolotl DNA have been 'shotgun' cloned into a bacterial plasmid and cloned in E. coli. Using the major repeat class as probe clones have been isolated which contain repetitive DNA. Several such clones have been analysed in some detail. Apart from allowing a more detailed analysis of the genome such clones may be used as probes for future comparative studies of the evolution of species in the genus Ambystoma.

The analysis of the axolotl genome is, of course, far from complete. A few of the more obvious future studies would be to examine reassociation products of DNA fragments of varying initial size in the electron microscope to complement the interspersion analysis already performed. Similarly, since the general organisation of the genome has to some extent been established, the fine organisation of repeated sequences could now be profitably analysed using large genomic inserts in λ or cosmid vectors. The clones already described would provide useful starting probes for such an investigation, such studies would help to determine the detailed sequence surroundings of those clones already described.

In the medium to long term the goal will be to use the present studies as a basis for comparative analyses within the Ambystomatids. The

conservation of repeated sequences between species can be analysed in a manner similar to that of previous work on the Plethodontids, cereals and sea urchins (see earlier discussions).

While the debate about the function of repeated sequences continues it is becoming increasingly important to analyse the larger genomes where questions about 'selfish', 'parasitic', 'ignorant', 'nucleotypic' DNA are perhaps of even more importance to our understanding of the relationship between sequence organisation and gene expression than in those smaller genomes which are more amenable to study, and which provide a large proportion of our present knowledge. The ideas and methods which have been stimulated by studies on small genomes should now be transferred to the larger genomes where the opportunity for confirmation and further discovery is much greater. The complexity of large genomes should not be viewed as an obstacle to investigations on the nature and evolution of repetitive DNA.

APPENDIX I

PHYLOGENY OF THE AMBYSTOMATIDAE

The Urodeles can be divided into four sub-orders (see Table AI.1), which can be further divided into eight families. Several independent methods of investigation suggest that the Cryptobranchoidea represent the most primitive form among the Urodeles. The karyotypes of the Cryptobranchoidea are similar to the most primitive forms of the Anura and Apoda (the two other orders of Amphibia) and are similar to various aquatic vertebrates and some reptiles (Morescalchi, 1980). The osteology of the Urodeles suggests a common ancestor of hynobiid stock (Tihen, 1958). The Hynobiids are present throughout the Palearctic region and a consistent phylogeny can be demonstrated for the families based on radiation from a hynobiid or hynobiid-like stock.

The relevant features of the phylogenetic relationships can be summarised as follows. Based on an extensive study of the karyotypes of the Urodeles Morescalchi (1980) suggests that the present day Cryptobranchoidea have retained most of the features of the ancestral stock. The Proteoidea show a more developed karyotype, lacking the large numbers of mini-chromosomes characteristic of the Cryptobranchoidea. The karyotype is, however, still asymmetric i.e. exhibiting both metacentric and acrocentric chromosomes. The Ambystomatoidea and Salamandroidea are similar in that their karyotypes are arranged as metacentrics with a fairly regular gradation of size. The Salamandroidea with 24 chromosomes could be viewed as descending from an Ambystomatoid stock, however although the phylogenetic distance between the two sub-orders is not great, current opinion is that the sub-orders were derived independently from a hynobiid stock (Morescalchi, 1980). The parallel development of the karyotype may represent a general evolutionary trend. It is worth noting that within the Cryptobranchoidea

the forms that are possibly the most differentiated have developed towards a symmetrical karyotype e.g. P. retardatus (2n = 40). Study of the karyotype has therefore allowed a general ordering of the relationship of the various sub-orders.

Tihen (1958) has analysed the phylogeny of the sub-order Ambystomatoidea in some detail. Based on osteological data he suggests the following phylogeny. As suggested above the Ambystomatoidea were probably derived from a hynobiid or pre-hynobiid stock. Early in the development of the Ambystomatoidea the separation into the three families occurred (see Table A I.1). By this time however the karyotype was probably reduced to 2n = 28 representative of the generalised forms of the Ambystomatoidea.

The Plethodontidae and Amphiumidae probably developed from an Ambystomatoid stock. The differentiation may have been occuring as the ambystomatids entered the Americas from Asia at the end of the Mesozoic period, 65 million years (myr) ago.

Early in the development of the Ambystomatoidea the migration branched (Table A I.2). This branching must have occurred after the development of those basic characteristics which separate the three families of the Ambystomatoidea. One branch became isolated west of what is now the Rocky Mountains, probably as a result of the Laramide orogeny. It is believed that the Dicamptodontinae and Rhyacotritoninae are derived from this western branch.

The eastern branch reached the general vicinity of the Great Lakes from which a major dispersal ensued. During the migration of the eastern branch the osteological characteristics of the Ambystomatinae developed. The main movement of the dispersal from the Lakes area was southward. The climate at this time prevented any major northward migration while the Central Plains were too arid for a major move west. It is believed that an ancestral form of Rhyacosiredon entered Mexico during the Tertiary period. Migration further south was prevented because at that time South America was

TABLE AI.1

CLASSIFICATION OF THE URODELES

SUB-ORDER	FAMILY
CRYPTOBRANCHOIDEA	HYNOBIIDAE
	CRYPTOBRANCHIDAE
AMBYSTOMATOIDEA	AMBYSTOMATIDAE *
	PLETHODONTIDAE
	AMPHIUMIDAE
SALAMANDROIDEA	SALAMANDRIDAE
PROTEOIDEA	PROTEIDAE
	SIRENIDAE

* SEE TABLE 2.

TABLE AI.2

CLASSIFICATION OF THE AMBYSTOMATIDAE

SUB-ORDER : AMBYSTOMATOIDEA

FAMILY : AMBYSTOMATIDAE

	SUB-FAMILY	GENUS	SUBGENUS	SPECIES
1)	DICAMPTODONTINAE	AMBYSTOMATICHNUS	-	A.montanensis
		DICAMPTODON	-	D.ensatus
2)	RHYACOTRITONINAE	RHYACOTRITON	-	R.olympicus
3)	AMBYSTOMATINAE	RHYACOSIREDON	-	R.altamirani
				R.leorae
				R.rivularis
				R.zempoalensis
		AMBYSTOMA	AMBYSTOMA	A.mexicanum
		MEXICANUM GROUP		A.lermaensis
				A.kansense
				A.tigrinum
				A.amblycephalum
				A.bonibypellum
				A.fluvinatum
				A.granulosum
		TIGRINUM GROUP		A.hibbardi
				A.lacustris
				A.ordinarium
				A.rosaceum
				A.subsalsum

TABLE A1.2 (cont.)

SUB-FAMILY	GENUS	SUBGENUS	SPECIES
AMBYSTOMATINAE	AMBYSTOMA	AMBYSTOMA	A.opacum
	OPACUM GROUP		A.talpoideum
			A.maculatum
			A.gracile
			A.macrodactylum
	MACULATUM GROUP		A.jeffersonianum*
			A.platineum*
			A.tremblayi*
			A.laterale*
		LINGUELAPSUS	A.annulatum
			A.cingulatum
			A.mabeei
			A.schmidti
			A.texanum
		BATHYSIREDON	B.dumerili

NOTE * A.jeffersonianum is a diploid species commonly found associated with the triploid (3n) A.platineum. Similarly A.laterale is a diploid species commonly associated with the triploid A.tremblayi. The four species are closely interrelated and are collectively known as the Jeffersonium complex (Uzzel Jr., 1964).

separated from North America by a strait further south than the present Panama isthmus, in the region of Venezuela and Colombia (Pomerol, 1982). Rhyacosiredon now represents the most southerly of the Ambystomatinae. A second migration into Mexico soon followed. During this migration it is thought that an adaptation to arid, unforested conditions took place and a northward dispersal from Mexico is postulated to have occurred. This is based on the appearance of fossil species of the mexicanum and tigrinum groups in middle and upper Pliocene deposits (2-4 myr ago) in the present day Kansas region (see Fig. A I.1). These fossils of A. kansense and A. hibbardi respectively are not found in earlier deposits although the A. mexicanum - A. tigrinum - Bathysiredon complex was well established in Mexico by this time.

At the same time as the development of the mexicanum - tigrinum group in Mexico further differentiation from the eastern stock took place in the Gulf region of the present day USA. The basic stock was differentiated from the mexicanum - tigrinum group. The sub-genus Linguelapsus diverged from this stock during the lower - middle Pliocene (3 - 5 myr ago). From this stock the opacum - maculatum ancestor also developed. The opacum group is thought to have developed from an isolated population of the maculatum stock in the south-east USA before the subsequent radiation of the maculatum group. Following the development of a more humid climate two northward migrations brought A. gracile and subsequently A. macrodactylum to the Pacific coast via a northerly route. The present maculatum group is thus represented in two ranges, a western range (A. opacum and A. macrodactylum) and an eastern range (A. maculatum and the A. jeffersonianum complex. Several of the current species distributions are available in the Catalogue of American Amphibians and Reptiles. These have been redrawn in Fig. A I.1.

separated from North America by a strait further south than the present Panama isthmus, in the region of Venezuela and Colombia (Pomerol, 1982). Rhyacosiredon now represents the most southerly of the Ambystomatinae. A second migration into Mexico soon followed. During this migration it is thought that an adaptation to arid, unforested conditions took place and a northward dispersal from Mexico is postulated to have occurred. This is based on the appearance of fossil species of the mexicanum and tigrinum groups in middle and upper Pliocene deposits (2-4 myr ago) in the present day Kansas region (see Fig. A I.1). These fossils of A. kansense and A. hibbardi respectively are not found in earlier deposits although the A. mexicanum - A. tigrinum - Bathysiredon complex was well established in Mexico by this time.

At the same time as the development of the mexicanum - tigrinum group in Mexico further differentiation from the eastern stock took place in the Gulf region of the present day USA. The basic stock was differentiated from the mexicanum - tigrinum group. The sub-genus Linguelapsus diverged from this stock during the lower - middle Pliocene (3 - 5 myr ago). From this stock the opacum - maculatum ancestor also developed. The opacum group is thought to have developed from an isolated population of the maculatum stock in the south-east USA before the subsequent radiation of the maculatum group. Following the development of a more humid climate two northward migrations brought A. gracile and subsequently A. macrodactylum to the Pacific coast via a northerly route. The present maculatum group is thus represented in two ranges, a western range (A. opacum and A. macrodactylum) and an eastern range (A. maculatum and the A. jeffersonianum complex. Several of the current species distributions are available in the Catalogue of American Amphibians and Reptiles. These have been redrawn in Fig. A I.1.

LEGEND TO FIG. A 1.1.

GEOGRAPHICAL DISTRIBUTION OF AMBYSTOMA SPECIES

Figure A 1.1 a - p shows the current species distributions for those species listed in the Catalogue of American Amphibians and Reptiles.

On each map the sites marked by a star and a number represent Tertiary fossil locations; 1, A. minshalli : Miocene - Pliocene transition; 2, A. minshalli : Mid-Upper Miocene; 3, A. kansense : Lower-Mid Pliocene; 4, A. hibbardi : Upper Pliocene.

Each map represents the location of one species of Ambystoma. The heavy shading represents the specific species location, the lighter shading represents the complete range for the genus.

a. A. mexicanum

b. A. tigrinum
   1. A.t. tigrinum
   2. A.t. mavortium
   3. A.t. nebulosum
   4. A.t. californiense
   5. A.t. melanoisticum
   6. A.t. velasei
   7. A.t. diaboli

c. A.ordinarium

d. A. rosaceum

e. A. jeffersonianum

f. A. laterale

g. A. platineum

h. A. tremblayi

i. A. talpoideum

j. A. annulatum

k. A. cingulatum

l. A. texanum

m. A. maculatum

n. A. gracile

o. A. macrodactylum

p. A. opacum

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

APPENDIX II

REASSOCIATION THEORY

A complete glossary of the terms involved in reassociation kinetic work can be found in Britten et al (1973). Here are listed terms most frequently used in this discussion.

1. DENATURATION : The disruption of a double helix into its complementary single strands, usually by heating, but also by treatment with acid or alkali. The temperature at which 50% of the nucleotides are denatured is known as the Tm.
2. HYPERCHROMICITY : The increase in ultraviolet absorption (u.v.) that occurs as double strands are converted to single strands.
3. HYPOCHROMICITY : The fall in u.v. absorption that occurs as double strands are formed from single strands.
4. RENATURATION : The formation of base paired structures, usually between two complementary strands, but potentially within one strand, if a sufficiently complementary region is present.
5. REASSOCIATION : The process of formation of double stranded molecules between two strands which are separate and at least one of which is free to diffuse.
6. CRITERION : The difference between the temperature of incubation (T_i) and the melting temperature in the reassociation solution of perfectly base paired DNA of the fragment size and %G+C under study.
7. COT. : The product of the DNA concentration (in moles nucleotides per litre $^{-1}$) and time (secs.) Conveniently expressed as : $\frac{\text{Absorbance at 260 nm (denatured) x hrs.}}{2}$. Alternatively a Cot of 1 is achieved when a DNA solution of 83 μg/ml is reassociated for 1 hour.

8. EQUIVALENT COT (E.Cot) : The term used when a correction factor is applied for salt concentration. A table based on an empirical formula is given on p. 373 Britten et al, (1973). The T_i must also be altered to allow for the change in criterion from that equivalent to $60^{o}C$ in 0.12M NaPB pH6.8 (a).

9. SEQUENCE COMPLEXITY : The total length of different DNA sequences, in nucleotide pairs.

10. KINETIC COMPLEXITY : The sequence complexity as calculated from the results of a measurement of the rate of reassociation of a DNA sample

$$G_i = K_i \times \text{correction factors} \times G_S/K_S$$

where G_i = kinetic complexity of sample i

K_i = observed rate constant of sample i

G_S = kinetic complexity of kinetic standard S

K_S = observed rate constant of kinetic standard S.

11. INTERSPERSION : The occurrence of recognisably distinct types of DNA sequences adjacent to each other, in an alternating but not necessarily regular pattern.

(a) NaPB pH6.8 is an equimolar solution of Na_2HPo_4 and Na_2Po_4. The molarity refers to the phosphate ion concentration.

REASSOCIATION

Early observations on the denaturation and renaturation of DNA were reviewed by Marmur et al (1963) and extended by Wetmur and Davidson (1968) and Britten et al (1973).

Briefly, initial observations on DNA RENATURATION indicated that a second order reaction was involved (Marmur et al,1963). Thus the rate limiting step is the collision of two complementary sequences which allow the base pairing of one or a small number of nucleotides. This step is known as the NUCLEATION EVENT. The subsequent base pairing along the rest of the overlapping complementary regions is fast with respect to nucleation.

The reaction can be expressed as :

$\frac{dc}{dt} = -kC^2$ where C is the concentration of nucleotides remaining single stranded at time t.

Integrating with $C = C_o$ at $t = 0$ gives :

$\frac{C_o}{C} = 1 + kCot$ where C_o is the concentration of nucleotides single stranded at $t = 0$ (t_o).

This can be rearranged to give :

$$\frac{C}{C_o} = \frac{1}{1 + kCot}$$

When $\frac{C}{C_o} = 0.5$, at half reaction,

$$Cot\ \tfrac{1}{2} = \frac{1}{k}.$$

Therefore the reassociation of any given DNA of known SEQUENCE COMPLEXITY can be described by its RATE CONSTANT, k, or by its $Cot\frac{1}{2}$. It should be evident also that there is a reciprocity between initial concentration and time in their relation to k. REASSOCIATION data may be expressed as $\frac{C_o}{C}$ vs Cot, giving k as the gradient of a straight line, or as the more familiar "Cot plot" (Britten and Kohne, 1968) of $\frac{C}{C_o}$ vs Cot. The point of 50% reaction is easily read from the plot and expressed as $Cot\frac{1}{2}$ or k.

RATE OF REASSOCIATION

1. METHODS OF MEASUREMENT

Marmur et al (1963) suggest several ways in which the extent of reassociation may be measured. The most commonly used are i) optical measurements on HYPOCHROMICITY, ii) digestion of reassociated products with S_1 nuclease, iii) Hydroxylapatite (HAP) chromatography.

i) Optical measurements

The earliest measurements on reassociation noted that denaturation by heating, followed by slow cooling caused DNA to become double stranded, the duplex regions of which had characteristics upon remelting similar to those of native DNA (Marmur, Rownd and Schildkrant, 1963). It was observed that the extent of reassociation could be monitored by noting the hypochromicity with time. In effect the increase in the duplex fraction was being measured.

Current common practice is to denature the DNA sample, shift the temperature rapidly to the temperature of incubation and monitor the decrease in optical density with time. (It must be noted however that there is an instantaneous decrease in optical density as DNA is cooled. This is termed the COLLAPSE HYPERCHROMICITY and is due to random stacking and base pair interactions within single stranded DNA. The effect is small, usually 2 - 5%).

An alternative method is to remove aliquots from a reassociation mix and melt the reassociated products. The HYPERCHROMICITY is a direct measure of the base pairing which has occurred.

ii) S_1 nuclease digestion of reassociated products

The single strand specific nuclease from Aspergillus oryzae (S_1 nuclease) will degrade single stranded DNA or RNA into 5' mononucleotides

in the presence of Zn^{2+} ions, but will not degrade double stranded DNA or DNA : RNA hybrids (Vogt, 1973). By removing aliquots from a reassociation mix and digesting the DNA with S_1 nuclease until all of the single strands are removed a measure of the SEQUENCE COMPLEXITY which has entered duplex DNA can be obtained.

iii) Hydroxylapatite (HAP) chromatography

Hydroxylapatite is a calcium phosphate complex which is capable of binding both double and single stranded DNAs. (Bernardi, 1965; Miyazawa and Thomas, 1965; Martinson, 1973 a + b; Britten et al. 1973). Single and double stranded DNAs bind to HAP, but with differing affinities, adsorption being controlled by phosphate ion concentration (see Martinson, 1973b for a discussion of the phase transitions of DNA - HAP). At 60°C and 0.12M NaPB pH6.8 single stranded DNA does not bind to HAP (Kohne and Britten, 1971; Martinson, 1973b). Therefore duplex DNA can be selectively adsorbed and subsequently eluted. Elution can be by :

(a) raising the temperature of the HAP - DNA complex above the temperature required for strand separation so that DNA becomes single stranded and can be eluted;

(b) raising the phosphate ion concentration. This can be done as a linear gradient (Bernardi, 1965) or by a single step to high salt (0.4M NaPB pH6.8; Kohne and Britten, 1971).

The principal difference between HAP chromatography and optical or S_1 methods is that HAP separates molecules which contain duplexes from molecules which are totally single stranded. Therefore single stranded regions physically associated with duplexing regions score as duplex whereas such sequences score as single strands with the other two methods. Therefore the fraction of sequences remaining single stranded is underestimated when reassociation products are analysed by HAP chromatography.

2. FACTORS AFFECTING RATE OF REASSOCIATION

The rate of reassociation is dependent upon collision of two complementary sequences. Therefore factors which may affect such collisions will affect the observed rate of reassociation. These factors include renaturation temperature, fragment length, monovalent cation concentration, sequence complexity, repetition frequency, viscosity of the solution and steric inhibition effects. It is proposed to discuss each in turn briefly, dealing more fully with those that directly influence the conclusions to be drawn from the data presented.

i) RENATURATION TEMPERATURE

Early studies on renaturation indicated that the optimum temperature for reassociation was 20 - 30°C below the Tm of the particular DNA fragments in question (Marmur et al, 1963). Bonner et al (1973) confirmed this relationship and showed that the optimal rate was achieved at Tm - 23°C although the reduction in rate was less than 10% of the optimum in the range 16 - 31° below Tm.

Many of the factors which affect rate have an indirect effect on Tm, so that corrections need to be made to maintain equivalent criterion of reassociation if any factor is altered. Here the effect of each is briefly noted, along with corrections which can be applied to maintain criterion. The major effect of each factor on rate will be discussed separately (later).

ia) % G + C content

Marmur and Doty (1959, 1962) have shown that the G:C base pairs are more thermally stable than A:T base pairs and have indicated that for each 1% G+C, the Tm is raised by 0.41°C at neutral pH above that for a poly d(A:T) molecule.

ib) Cation concentration

Felsenfeld (1962) and Marmur and Doty (1962) have shown that the

stability of DNA is increased by masking of the negatively charged phosphate backbone. This can be achieved by an increase in ionic strength in solution. Reducing the ionic strength to $\frac{1}{10}$th the original concentration reduces Tm by 16.6°C, and the change is linear over a wide range of Na^+ ion concentrations.

Taken together the effect of GC content and cation concentration can be given as

$$Tm = 81.5 + 16.6 \ (\text{Log } [Na^+]) + 0.41 \ (\% \ G+C).$$

ic) Fragment length

Britten et al (1973) have introduced an empirical equation to observations on the effect of fragment length on Tm.

$Tn = Tm + \frac{B}{L}$ where Tn is the melting temperature of long DNA (native)

Tm is the melting temperature of DNA of length L

B is a constant.

for $[Na^+]$ between 0.05 and 0.5M,

$B = 300 + 2000 \times [Na^+]$ Britten et al (1973).

id) Organic solvents

In early studies a range of organic solvents were examined for their effects on Tm (Rice and Doty, 1957; Marmur and T'so, 1961; T'so et al, 1962 a + b; Levine et al, 1963). The compound most commonly used is formamide. The effect of formamide on Tm has been extensively analysed (McConnaughy, David and McCarthy, 1969) and it appears that for each 1% formamide in the renaturation the Tm falls by 0.72°C. Formamide is commonly added to DNA - RNA hybridisations to reduce the incubation temperatures and so reduce the RNA degradation during hybridisations. When HAP analysis followed by optical measurements of duplex DNA are made the common practice is to dilute the reassociated aliquot to 1%

formamide before loading and to monitor absorbence at 270 nm. Here nucleic acid absorbence is approximately 80% of that at 260 nm whereas absorbence due to the formamide is minimal, (formamide absorbs strongly at 260 nm, McConnaughy et al, 1969).

ie) Divergence

Bonner et al (1973) have determined that for every 1% mismatch between sequences in a reassociation event, the Tm is reduced by 1°C. In a complex reassociation mix such as occurs in the reassociation of higher eukaryotes a range of divergence among sequence families may exist. Thus it would not be possible to correct the incubation temperature to optimise the reassociation of all the repeat families. In such cases it should be stressed that the most rapidly reassociating sequences need not be the most highly repetitive (see for example the human Alu I family, Houck et al, 1979).

For most DNAs, with a % G+C content of around 50%, the optimum rate is achieved at 60°C in 0.18M Na^+. As single stranded DNA does not bind to HAP at 60°C in 0.12M NaPB (Britten et al, 1973; Martinson, 1973 a + b) it has become usual practice to use 60°C and 0.18M Na^+ as the standard reassociation conditions. Under this reassociation criterion sequences up to 25% diverged can reassociate (Bonner et al, 1973). By varying the temperature of reassociation sequences of varying homology can be renatured (see above, ie). However, it is important to remember the effects of rate when reactions are performed beyond the optimum regions, especially in the case of complex genomes.

ii) FRAGMENT LENGTH

There are two effects on rate, an indirect effect due to an effect on Tm (see above) which must be corrected for when reassociation criterion is determined, and secondly an effect directly related to the fragment length.

For a DNA in which sequence complexity is equivalent to genome size the observed rate of reassociation should increase with fragment length as nucleation, not subsequent base pairing along the length of the sequence, is rate limiting, Wetmur and Davidson (1968) showed however that the rate is proportional to the square root of the length. They attributed this retardation to an excluded volume effect. Kinberg-Calhoun and Wetmur (1981) have recently demonstrated using DNAs of defined length and sequence composition that such an excluded volume effect actually exists. Smith et al (1975) and Britten and Davidson (1976) have noted a similar steric effect which retards reassociation late in a reaction.

Thus for a simple genome,

$$k \propto \frac{L}{G}^{0.5}$$ where k = reassociation rate constant
L = fragment length
G = sequence complexity.

For a genome in which sequence repetition occurs the observed effect of length depends upon the method of analysis. Where sequences of varying repetition frequency are interspersed, and for fragment lengths which are larger than the average family sequence length the most highly repetitive element on a fragment will, on average, reassociate first. Upon assay by HAP the whole fragment would score as duplex whereas by optical or S_1 nuclease methods only the region of duplex DNA would score as duplex. Thus the observed rate will depend on the overlap size if optical and S_1 methods are used, and on the fragment length if HAP is used to assay the reassociation products.

As the overlap size varies depending upon the family size and interspersion pattern, which differ for individual repetitive families and for different organisms, predicting general effects on rate is not feasible. Where such effects need to be measured empirical relationships should be obtained by experiment on the species and repetitive families in question.

In reactions in which a small amount of labelled DNA (tracer) is

added to a large excess of unlabelled (driver) DNA the rate of reassociation of the tracer is initially dependent upon its length relative to the driver DNA. Data by Wetmur (1971) and Hutton and Wetmur (1973) have shown that DNA strands of different lengths present in stoichiometric amounts reassociate with a rate constant proportional to the square root of the length of the shorter strand. However when the driver is present in a vast excess the rate constants of the two strands cannot be predicted by reference to these earlier works. Hinnebusch et al (1978) and Chamberlin et al (1978) have analysed in detail the effects of reassociating tracer DNA of various lengths with excess driver DNA.

Chamberlin et al (1978) suggest that in long tracer - short driver reactions the observed rate constant for the tracer reaction increases proportionately with tracer length, the relationship being given as

$$K_T = K_D \frac{\bar{L}_T}{\bar{L}_D}$$

where K_T is the tracer/driver rate constant
K_D is the rate constant for the driver self reaction,
and $\bar{L}_T$ and $\bar{L}_D$ are the mean single strand lengths of tracer and driver DNAs respectively.

They further suggest that this relationship does not hold when the driver length ($\bar{L}_D$) is greater than the tracer length ($\bar{L}_T$) and prefer the following relationship to fit the data.

$$K_T = K_S \left(\frac{\bar{L}_T}{\bar{L}_D} \right)^{0.5}$$

where K_T, $\bar{L}_T$ and $\bar{L}_D$ are as above,
K_S is the rate constant for a tracer of length $\bar{L}_T$ with driver DNA of the same size.

Chamberlin et al emphasise that these formulae are empirical and are designed for the purpose of calculating expected tracer rate constants within the accuracy of most experiments. They suggest that a more complex

fraction of $\bar{L}_D$ v $\bar{L}_T$ can be described for the complete range of $\bar{L}_D > \bar{L}_T$ to $\bar{L}_T > \bar{L}_D$ but that "no significant gain in accuracy or understanding is thus achieved".

Therefore in situations where, for example, a nick translated DNA is to be used as a tracer the mean single strand size should be known so that an accurate description of the reaction, based on tracer Cot values, can be determined.

Hinnebusch et al (1978) have produced a formula which can be used to relate the tracer rate constant to any combination of tracer or driver lengths. Their formula, based on empirical observations on the reassociation of simple DNAs of carefully defined lengths is given as

$$K_T = a.\bar{L}_T \left(\frac{1}{\bar{L}_T^{0.55}} + \frac{1}{\bar{L}_D^{0.55}} \right)$$

where K_T, $\bar{L}_T$, $\bar{L}_D$ have the meanings already noted above, and a is a constant.

when $\bar{L}_T = \bar{L}_D$

$$K_T = a(\bar{L}_T)^{0.45}$$

In the discussion on tracer - driver reactions so far the conclusions have been confined to the reassociation of simple DNAs. Hinnebusch et al (1978) have addressed the problem of whether, in a tracer - driver reassociation involving repetitive sequences, the value of Cot to which the reassociation is carried should be adjusted with changing tracer length so as to include in each reaction only those tracer fragments containing a repeated sequence.

They suggest that the tracer reactions involving repeated sequences will not change appreciably with changing tracer length. However, they do suggest that when the tracer length is very large compared to the driver length then reassociation of tracer single copy DNA may interfere with the interpretation of the results. This would lead to an overestimate of the fraction of the genome that is organised as interspersed repeated and unique sequences. Their data suggest that if, when $\bar{L}_T = \bar{L}_D$ the $Cot\frac{1}{2}$ values for repeated and unique sequences are not separated by at least

4 decades of Log Cot and total DNA is used as driver, then tracer single copy DNA reassociation will be appreciable at driver repetitive Cot values, when $\bar{L}_T \gg \bar{L}_D$. The practical implications of these problems have been discussed further for specific instances by Murray et al (1978) and Cullis (1981).

To resolve this problem suitable Cot points for the reaction should be chosen or the driver DNA should be prefractionated to remove single copy DNA not adjacent to repeated sequences.

iii) DIVERGENCE

For a genome which contains repetitive DNA the effect of sequence divergence due to mutation is to reduce the rate of reassociation of the sequence family. Bonner et al (1973) have estimated the effect on the rate of reassociation of sequence divergence. By means of controlled deamination of DNA, mismatches were induced. The rate of reassociation of DNA was found to be reduced by twofold for every 10% mismatching. The effect is probably due to inhibition of stable nucleation sites.

iv) STERIC EFFECTS

As noted above Wetmur and co-workers have shown that the effects of length on rate are not straightforward and have put forward an exclusion effect hypothesis to explain the retardation in rate. They have shown that for the reassociation of randomly sheared DNA the reassociation of the single strand "tails" is impeded, most probably due to steric effects, so reducing the overall rate. Britten and co-workers have analysed this problem with particular reference to the observed rate late in the reassociation of simple DNAs (Smith et al, 1975; Britten and Davidson, 1976). They have shown, both for HAP and S_1 measured reactions, that there is a definite inhibition effect late in reassociation. Britten and Davidson (1976) have shown by computer simulations that measured HAP

kinetics are noticeably different from calculated kinetics. The observed kinetics are initially faster than predicted and slower than predicted towards the end of the reassociation. This is because free single strands can reassociate with single strand tails, and single strand tails with each other so that the overall reaction has the form,

$$\frac{dc}{dt} = -kC^2 - kC\,(S-C)$$ where S-C is the concentration of single strands physically linked to duplex DNA.

therefore the observed rate should be faster than second order. However the variation in fragment length in experimental situations obscures this distinction because larger molecules reassociate faster causing the mean size of single strands to fall during the course of a reassociation experiment. The observed terminal reduction in rate is not however totally due to the reduced mean size and a true inhibition does occur late in a reassociation. This can be seen from studies of optical or S_1 nuclease monitored reassociations. Here both free and bound single strands are scored as unreassociated. Any reduction in rate late in reassociation must therefore be due to an inhibition effect. Smith et al (1975) have shown that the rate equation which best fits the experimental data can be written as

$$\frac{S}{C_o} = \frac{1}{(1 + KCot)^{0.45}}$$ where S = fraction of the sequence complexity which is single stranded at time = t.

However, given that the overlap between two randomly fragmented sequences is 55% (see above) the ideal equation governing this type of reassociation should be

$$\frac{S}{C_o} = \frac{1}{(1 + KCot)^{0.55}}$$

Observed reactions begin with the predicted initial rate but the rate falls until the best fit value for n = 0.45.

In the reassociation reaction therefore an inhibition effect occurs

which becomes observable as the reaction approaches completion. The extent of this inhibition has been estimated to reduce the terminal rate by 2-4fold (Britten and Davidson, 1976). It would thus appear that the region close to the end of a duplex sequence is not available for reassociation, so reducing the effective number of nucleation sites, and the observed rate.

The above observations are limited to simple sequences, e.g. where there is no sequence repetition. No measurements have been made for the case where the tail constitutes the majority of the fragment length e.g. in long period interspersion patterns. It is possible that the reduction in rate will not be of the same order as the probability of nucleation within the region affected by inhibition effects will be lowered as tail size increases.

v) IONIC STRENGTH

The rate of nucleation can be increased by neutralisation of the negatively charged phosphate groups. This can be effected by an increase in ionic strength. A concomitant effect is the stabilisation of DNA duplexes which has an effect on Tm (see above). The effect of monovalent cation concentration on rate (after correction to maintain criterion) has been estimated for a range of concentrations by Britten et al (1973) at Tm - 25°C. They have developed an empirical formula to describe the relationship

$$\text{Rate} \propto a^{(0.24/a)^{0.42}}$$ where a = monovalent cation concentration.

The rate is also related to the ionic radius of the cation (Zimmer and Fenner, 1962; cited in Marmur et al, 1963). However as Sodium is the usual cation used the effects of other cations shall not be detailed except to note that some act by raising Tm by neutralising phosphate groups

(e.g. Mg^{2+}, Mn^{2+}, Co^{2+}, Zn^{2+}) whereas others reduce Tm by displacing H bonds by their coordination to electron donor groups (e.g. Cr^{2+}, Cd^{2+}, Pb^{2+}) (Eichhorn 1962). It is useful in reassociations therefore to include a chelating agent in small concentrations or to pass the DNA over an ion exchange column before use in order to minimise the effects of such ions on rate. (See also Sissoeff et al, 1976, for a review of DNA-metalion interactions and effects).

vi) % G+C CONTENT

Wetmur and Davidson (1968) have shown by measuring rate constants of DNA of known % G+C and complexity that the rate increases slightly with increasing % G+C content. They propose that the relative rate constants for the range 34 - 64% G+C are 0.7 - 1.3 of that for 50% G+C. The relationship outside of these limits is unclear. (See Fig. A II.1).

vii) pH

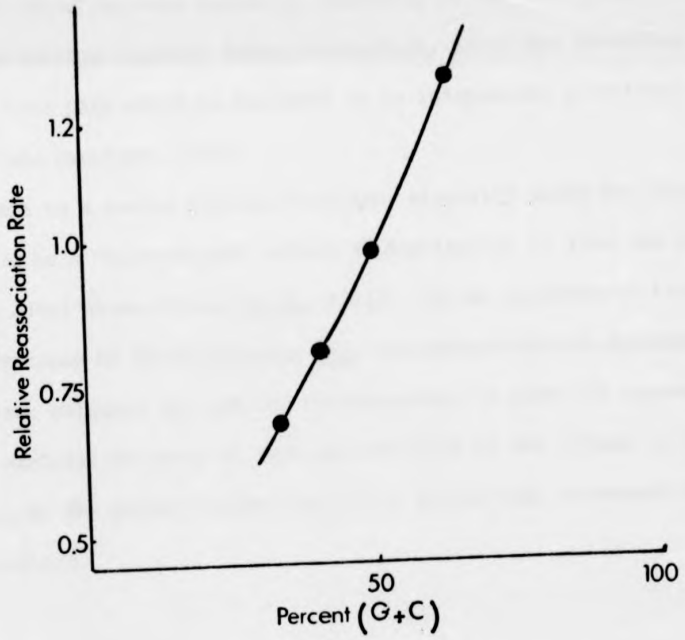
Dove and Davidson (1962) have shown that the stability of DNA is dependent upon pH. At 20°C DNA denatures below pH 2.7 and above pH 12. The denaturation is dependent on ionic strength, temperature and base composition. Low temperature and high G+C content tend to make DNA resistant to acid denaturation, whereas high pH denaturation is dependent only on % G+C.

At pH's between 5.5 and 8.5 the Tm is not greatly affected (T'so et al, 1962a). Therefore no correction needs to be made to the temperature of the reassociation which is usually carried out at pH 6.8 - 7. Similarly within the pH range 5 - 9 the rate of reassociation is not affected. The rate is reduced markedly at pH above 9 (Wetmur and Davidson, 1968) by an extent greater than can be accounted for by the effect of pH on Tm.

LEGEND TO FIG. A II.1

RELATIVE REASSOCIATION RATE : DEPENDENCE UPON %G+C

Curve based on the data of Wetmur and Davidson (1968) relating rate of reassociation to %G+C.


Relative Reassociation Rate
1.2
1.0
0.75
0.5
50
100
Percent (G+C)

viii) VISCOSITY

The effects of viscosity are included for completeness as the reassociations discussed in this thesis do not require corrections for viscosity changes. Subiriana and Doty (1966) demonstrated that the rate of renaturation decreased with increasing solvent viscosity. Wetmur and Davidson (1968) investigated this effect in some detail. As well as an overall effect on rate there was also an effect on Tm, usually due to chemical effects of the added substance.

The effect on rate caused by viscosity is due to a retardation of the base pairing reaction during nucleation, after the formation of the initial base pair which is believed to be independent of solvent viscosity (Wetmur and Davidson, 1968).

There is a second effect of solvent viscosity which has been described as a "macroscopic" effect to distinguish it from the microscopic effects noted above (Chang et al, 1974). In the presence of high concentrations of added polymers e.g. tetraethylammonium chloride (TEACL) or dextran sulphate the rate of reassociation is markedly increased, the simple explanation being an exclusion of part of the volume of the solution by the polymers resulting in an effectively increased DNA concentration.

REPETITIVE SEQUENCES

Most of the above discussion has been specifically related to genomes with one major sequence class. Eukaryotes have been shown to have sequences repeated many times within the genome (Britten and Kohne, 1968).

Suppose a repetitive sequence occurs F times per haploid genome and in total occupies a fraction, α, of the genome. The complexity, N of this sequence family (assuming in this simple case no divergence) is given by $N = \frac{\alpha G}{F}$ where G = genome size in base pairs.

This reduces to N = G where the genome is entirely non-repetitive.

When the genome is sheared to a size less than or equal to the size of the repetitive sequence then the concentration, Cr, for the repetitive sequence is $Cr = \frac{F\ Co}{G}$ where Co is the total concentration of nucleotides.

$$\frac{Cr(ss)}{Cr} = \frac{1}{1+\left(kr\ \frac{F}{G}\right)Cot}$$

It can be seen that the observed rate constant will be proportional to F, the repetition frequency, and inversely proportional to G, the genome size. Therefore F can be calculated by

$F = \frac{kr}{Ksc}$ where Ksc is the rate constant for a single copy fraction of the same genome.

The overall reassociation for a complex genome is the sum of the reassociation of the repetition classes which make up the genome.

$$\frac{C}{C_o} = B + \frac{\alpha_1}{1+k_1Cot} + \frac{\alpha_2}{1+k_2Cot} + \frac{\alpha_3}{1+k_3Cot} \cdots\cdots\cdots$$

Where B is the fraction of DNA remaining unreassociated at the termination of the reaction (and usually represents degradation of DNA during reassociation) and each component α_i of the genome reassociates with the observed rate constant k_i.

The usual method of analysing such data is to use a non linear least squares fit with several components. This is conveniently achieved using a computer.

It is important to note, as has been mentioned earlier, that when complex genomes are analysed using hydroxylapatite the DNA fragments become scored as duplex at the rate appropriate for the most rapidly reassociating sequence (usually the most highly repetitive sequence, but not necessarily so, see the section on effect of divergence on rate,

This reduces to F = G where the genome is entirely non-repetitive.

When the genome is sheared to a size less than or equal to the size of the repetitive sequence then the concentration, Cr, for the repetitive sequence is $Cr = \frac{F\ Co}{G}$ where Co is the total concentration of nucleotides.

$$\frac{Cr(ss)}{Cr} = \frac{1}{1+\left(kr\ \frac{F}{G}\right)Cot}$$

It can be seen that the observed rate constant will be proportional to F, the repetition frequency, and inversely proportional to G, the genome size. Therefore F can be calculated by

$F = \frac{kr}{Ksc}$ where Ksc is the rate constant for a single copy fraction of the same genome.

The overall reassociation for a complex genome is the sum of the reassociation of the repetition classes which make up the genome.

$$\frac{C}{C_o} = B + \frac{\alpha_1}{1+k_1Cot} + \frac{\alpha_2}{1+k_2Cot} + \frac{\alpha_3}{1+k_3Cot} \cdots\cdots\cdots$$

Where B is the fraction of DNA remaining unreassociated at the termination of the reaction (and usually represents degradation of DNA during reassociation) and each component α_i of the genome reassociates with the observed rate constant k_i.

The usual method of analysing such data is to use a non linear least squares fit with several components. This is conveniently achieved using a computer.

It is important to note, as has been mentioned earlier, that when complex genomes are analysed using hydroxylapatite the DNA fragments become scored as duplex at the rate appropriate for the most rapidly reassociating sequence (usually the most highly repetitive sequence, but not necessarily so, see the section on effect of divergence on rate,

earlier.) Where interspersion of repetitive classes occurs the more highly repetitive fraction will therefore be overestimated and the lower repetitive fraction(s) will be underestimated. The magnitude of this effect depends on the fragment length and on the spacing of the interspersed sequences.

When such analyses are performed, in many cases there does not appear to be sharp transitions from one repetitive class to the next. (A single kinetic component can occupy only two decades of Cot between 10 and 90% of the reaction). The normal practice is to resolve the data on the basis of the best fit using the minimum number of second order components, while remembering that such methods merely reduce the data to manageable numerical averages.

The observed rate constants for each repetition frequency class need to be adjusted before final values for repetition frequency and complexity can be obtained. This is implicit in the rate equation above for complex genomes. A sequence family contained in a fraction α_i of the genome of size G base pairs is effectively diluted by the presence of $G-\alpha_i$ other sequences. Hence the true rate, kpure, can be obtained from

$$\text{kpure} = \frac{K_i}{\alpha_i}$$

where K_i is the observed rate constant.

The complexity can then be calculated with reference to a kinetic standard.

INTERSPERSION OF REPETITIVE SEQUENCE CLASSES

When moderately sheared eukaryotic DNA (2 - 5 kilobases) is reassociated to a point at which most repetitive sequences have renatured and the duplex containing DNA, recovered by HAP chromatography, is denatured, the hyperchromicity exhibited by the DNA suggests than on average only a small proportion of the recovered DNA is double stranded. This implies that repetitive and non-repetitive sequences are arranged in the genome in an

interspersed fashion. The hyperchromicity measurements give an estimate of the length of the interspersed repeat, as follows :

(1) The duplex length (L) is the total fragment length (f) x the fraction in duplex (d) $L = D \times F$.

(2) D can be inferred from the hyperchromicity (H) and assumes that for each fragment length and observed Cot the hyperchromicity of the duplex after treatment with S_1 nuclease has been measured.

If H_D is the hyperchromicity of renatured Duplex DNA and 4% is the accepted hyperchromicity of single strands, then

$$D = (H - 0.04)/(H_D - 0.04)$$ Graham et al, (1974).

Recently however Moyzis et al (1981,b) have suggested that these formulae underestimate the actual repeat length and should be re-evaluated. It has already been noted that in randomly sheared DNA the results of reassociation yield reassociated molecules with on average only 50% of the DNA as duplex, the rest being single strand "tails". Britten and Davidson (1976) have indicated that in order for these "tails" to reassociated the reassociation must be allowed to proceed for several decades of Cot past the $Cot_{\frac{1}{2}}$ (observed) for the reassociation. This is not possible as subsequent reassociation of slower reassociating sequences interferes with the analysis. It is therefore possible that in many cases the unreassociated "tails" may be of similar sequence complexity to the duplex and hence an underestimate of true repeat length will be made.

It is more usual however in these types of experiment to use fairly long fragments of DNA. This tends to reduce, but not eliminate the problem noted above. In these cases reassociating repeated sequences would have a higher probability of being enclosed on either side by non repetitive sequences (assuming that short period interspersion occurs). These duplexes would form an X shaped structure (a Y form would appear if the duplex was at one end of the reassociation product). These structures can be visualised in the electron microscope and have been put forward as evidence

for short period interspersion. The hyperchromicity of such products would be predicted to fall as the fragment length (and hence the proportion of "tail" increases). From the observed hyperchromicities of reassociated products formed from various initial lengths a consensus estimate for the length of the duplex region can be obtained (Graham et al, 1974). However in most cases the duplexes are recovered after reassociation by hydroxylapatite chromatography. Several groups have noted that fractionation of large fragments on hydroxylapatite results in retention of noticeable amounts of reassociated products. (Thompson, 1976; Flavell and Smith, 1977; Kiper and Herzfeld, 1978; Kiper, 1978; Moyzis et al, 1981a). Moyzis et al (1981a)have demonstrated that in human, rat and hamster DNA the retained DNA is a non-random sample of the reassociated products comprising networks and hyperpolymers of closely interspersed repeated sequences. The eluted fraction contains the proportion of fragments with few repeats per fragment length, which do not enter into hyperpolymers. By performing hyperchromicity and electron microscopy studies on such a sample, and then extrapolating the results to include the entire repetitive fraction may be leading to a false impression of both the repeat length and pattern of interspersion.

These methods of analysis do not give any information on the length of interspersed non-repetitive sequences or on the relative proportions of the genome which may be organised in the various patterns of interspersion. A method of analysing this problem was introduced by Davidson et al (1973).

Briefly, tracer amounts of radiolabelled DNA of varying single strand sizes are mixed with a vast excess of short unlabelled driver DNA fragments. The DNA is denatured and allowed to reassociate until the repeated sequences have renatured. Aliquots are removed and completely single stranded DNA is separated from the duplex fraction by hydroxylapatite chromatography.

The analysis depends upon the following points. If repetitive sequences are interspersed with single copy sequences many short fragments

will not contain repetitive elements. As the fragment length is increased the fraction of fragments that do contain repetitive sequences will increase. When the fragment size is longer than the length of the intervening non-repeating element in a particular interspersion class, then all fragments will contain repetitive sequences, and further increases in fragment size will not increase the observed size of this interspersion class. When more than one pattern of interspersion is involved the analysis can be viewed as the sum of the effects of each and the effect on binding to HAP will be dependent upon the tracer fragment length. (See Fig. 4.2).

It is possible from such data to derive :

i) the length of the interspersed non-repetitive elements;

ii) the fraction of the genome occupied by both repetitive and non-repetitive elements;

iii) the fraction of the genome which exhibits interspersion.

Thus this type of analysis is extremely useful in an analysis of genome organisation.

To aid in analysing this type of experiment several groups have produced a theoretical framework from which detailed descriptions of genome organisation may be obtained using this type of experiment (Davidson et al, 1973; Graham et al, 1974; Cech et al, 1973; Cech and Hearst, 1976; Schmid and Deininger, 1975; Jeffreys et al, 1977; Moyzis et al, 1981 a; Robertson et al, 1981).

The models vary in complexity depending on the detail in which the various variables are considered. However each provides statistical fits to the data to achieve estimates for the average lengths of the non-repeated sequences and the fraction of the genome occupied by the various interspersion classes. In practice the predictive ability of the models relies upon the accuracy of the data and upon the application of various corrections to the experimental points before any conclusions can be drawn.

The most detailed analysis to date has been prepared by Robertson et al (1981) and Moyzis et al (1981 a). In these papers is a detailed discussion of the various corrections which need to be made to the experimental data, and an assessment of past descriptive models. Their mathematical analysis is similar to, but generally more comprehensive than, other published models and while many of their experimental suggestions and corrections are already in the literature they provide a comprehensive summary of both theory and practice. In the process they suggest a novel alternative view of DNA sequence organisation (see main text).

COMPUTER PROGRAM FOR THE ANALYSIS OF REASSOCIATION KINETICS

The computer program used to analyse the reassociation kinetic data was that of Kells and Straus (1977). The program was modified for use on the Burroughs B6700 computer at the University of Warwick Computer Unit.

The program can be used in two forms depending on whether data from hydroxylapatite or S_1 nuclease assays was to be analysed.
The relevant difference concerns line 40400.

For hydroxylapatite data the formula used is :

$$FF(L,K) = P(J-1)/\lceil 1.0000 + P(J) * COT(K) \rceil$$

For S_1 nuclease data the formula used is :

$$FF(L,K) = (P(J-1)/\lceil (1.0000 + P(J) * COT(K)) \rceil **0.45$$

The listing shown on the following pages is that which would be used to analyse hydroxylapatite data (see line 40400).

```
#
L
100 $SET LINEINFO
200       EXTERNAL CRITER
300       COMMON /SCR/SER,SCRAT
400       COMMON COT(50),FF(5,50),CCO(50)
500       DIMENSION TIME(50),CO(50),CL(50),CT(50),Y(50),CCALC(50),TITLE(12),
600      * COTL(50),SCRAT(947),DIFF(11),SIGNS(11),P(11),SER(11)
700       DATA DIFF/11*0.01/,SIGNS/11*1.0/
800 C*          SIGNS=1.0 ENSURES ALL PARAMETERS WILL NOT CHANGE SIGNS FROM
900 C*          THOSE READ IN INITIALLY
1000 C* ***************************************************************
1100 C*
1200 C*       INPUT OF DATA AND REORDERING
1300 C*
1400  1000 READ(5,100,END=1111)NPARAM,TITLE,(P(I),I=1,NPARAM)
1500   100 FORMAT(I3,/,12A6,/,10F8.0)
1600 C*
1700 C*       NPARAM      NO. OF PARAMETERS TO BE USED IN SEARCH
1800 C*       TITLE       EXPERIMENT IDENTIFICATION
1900 C*        P          ESTIMATES OF PARAMETERS. PARAMETERS TO BE SUBMITTED I
2000 C*                  THE FOLLOWING ORDER:(TERMINAL FRACTION UNREASSOCIATED);
2100 C*                   (FRACTION CONTRIBUTED BY COMPONENT 1); (RATE OF
2200 C*                   REACTION OF COMPONENT 1); (FRACTION CONTRIB. BY
2300 C*                   COMPT. 2); (RATE OF COMPT. 2);ETC.,ETC.
2400       EPSPAR=0.00001
2500       EPSSSQ=0.0000001
2600       NOIT=35
2700 C*      EPSSSQ SSQ CONVERGENCE CRITERION. IF EPSSSQ=10**-K CALCULATION
2800 C*      STOPS IF(I-1)ST SSQ AGREES WITH (I'TH) SSQ TO K DECIMAL PLACES
2900 C*      EPSPAR PARAMETER CONVERGENCE CRITERION. IF EPSPAR=10**-K
3000 C*      CALCULATION STOPS IF ALL PARAMETERS FOR (I-1)ST ITERATION AGREE
3100 C*      WITH I'TH ITERATION TO K DECIMAL PLACES.
3200 C*      NOIT IS THE MAXIMUM NNO. OF ITERATIONS TO BE PERFORMED.
3300 C*
```

```
      DO 10 J=1,50
      READ(5,101) CCO(J),TIME(J),CO(J)
C*      LAST DATA CARD MUST BE 0.000 0.000 0.000
C*      CCO CONCENTRATION/INITIAL CONCENTRATION (% SINGLE STRANDED )
C*      TIME   IN HOURS
C*      CO  INITIAL CONCENTRATION OF DNA
  101 FORMAT(3F6.0)
      IF(CCO(J).EQ.TIME(J).AND.TIME(J).EQ.CO(J).AND.CO(J).EQ.0.0) GO TO
     *  11
      COT(J) = TIME(J)*CO(J)*0.5
      COTL(J) = ALOG10(COT(J))
   10 CONTINUE
   11 NPT = J-1
      DO 12 J=1,NPT
      KK = 1
      S = COTL(1)
      DO 13 K =2,NPT
      IF (COTL(K)-S) 14,13,13
   14 S = COTL(K)
      KK = K
   13 CONTINUE
      CL(J) = S
      CT(J) = COT(KK)
      Y(J) = CCO(KK)
      COTL(KK) = 100000
   12 CONTINUE
      DO 15 J =1,NPT
      COT(J)=CT(J)
      COTL(J)=CL(J)
      CCO(J)=Y(J)
   15 CONTINUE
```

```
C*
C*       ********************
C*       SEARCH FOR PARAMETERS
C*
        WRITE(6,500)TITLE
        CALL GAUS(CRITER,NPT,CCO,NPARAM,P,DIFF,SIGNS,EPSSSQ,EPSPAR,
       * NOIT,0.01,10.0,SUMSQ)
C*       ********************
C*
C*       DATA OUTPUT
C*
  500 FORMAT(1H1//5X,12A6//)
  501 FORMAT(2X,"COMPONENT",10X,"FRACTION",25X,"RATE",16X,"COMPLEXITY",
     *15X,"KPURE"//18X,"+/-STD.ERROR",17X,"+/- STD.ERROR",6X,"1.1E06*
     *FRACTION/RATE",7X,"RATE/FRACTION"//)
  502 FORMAT(6X,I1,4X,E11.4,1X,"+/- ",E10.3,5X,1P, E11.4,"+/-",E10.3,6X
     *1P, E11.4,12X,1P, E11.4//)
       WRITE (6,500) TITLE
       WRITE (6,511) P(1),SER(1)
  511 FORMAT(2X,35HTERMINAL FRACTION UNREASSOCIATED -   ,E11.4,"   +/- "
     *,E10.3,"  STD. ERROR"//)
       WRITE (6,501)
       NPAR1 = NPARAM - 1
       DO 16 J = 2,NPAR1,2
       L = J/2
       Q = 1100000.00*P(J)/P(J+1)
       QQ = P(J+1)/P(J)
       WRITE(6,502) L,P(J),SER(J),P(J+1),SER(J+1),Q,QQ
    16 CONTINUE
       CALL CRITER(P,CCALC,NPT,NPARAM)
       NCOMPT =NPAR1/2
  503 FORMAT(2X, 64H  C/C(0)        C/C(0)         COT         LOG(COT)
     *  DIFFERENCE  )
  504 FORMAT(2X, 22H   EXPT'L        CALC   //)
```

```
 505 FORMAT(2X, 90H   C/C(0)       C/C(0)        COT         LOG(COT)
     *  COMPT 1       COMPT 2     DIFFERENCE  )
 506 FORMAT(2X, 103H  C/C(0)       C/C(0)        COT         LOG(COT)
     *  COMPT 1       COMPT 2     COMPT 3     DIFFERENCE  )
 507 FORMAT(2X, 116H  C/C(0)       C/C(0)        COT         LOG(COT)
     *  COMPT 1       COMPT 2     COMPT 3     COMPT 4     DIFFERENCE  )
 508 FORMAT(2X, 129H  C/C(0)       C/C(0)        COT         LOG(COT)
     *  COMPT 1       COMPT 2     COMPT 3     COMPT 4     COMPT 5
     *DIFFERENCE  )
 509 FORMAT(2X,9(1P,E11.4,2X),1P,E11.4)
 510 FORMAT(1H0)
      WRITE (6,500) TITLE
      IF(NCOMPT.EQ.1)  WRITE(6,503)
      IF(NCOMPT.EQ.2)  WRITE(6,505)
      IF(NCOMPT.EQ.3)  WRITE(6,506)
      IF(NCOMPT.EQ.4)  WRITE(6,507)
      IF(NCOMPT.EQ.5)  WRITE(6,508)
      WRITE(6,504)
      L=0
      DO 19 J =1,NPT
      L = L + 1
      DIF  = CCALC(J) - CCO(J)
      IF (L.EQ.5)  L = 0
   17 IF (NCOMPT.EQ.1) WRITE(6,509)   CCO(J),CCALC(J),COT(J),COTL(J),DIF
      IF (NCOMPT.NE.1) WRITE(6,509)   CCO(J),CCALC(J),COT(J),COTL(J),
     *(FF(K,J),K=1,NCOMPT),DIF
      IF (L.EQ.0)  WRITE(6,510)
   19 CONTINUE
C*
C*      ************************
      GO TO 1000
 1111 STOP
        END
```

```
13200       SUBROUTINE GAUS(FOF,NOB,Y,NP,TH,DIFZ,SIGNS,EPS1,EPS2,MIT,FLAM,
13300      *FNU,SSQ)
13400 C*  GAUS-ORIG.PROG.-D.A.MEETER,ADAPTED-P.J.WOLFE,WISCONSIN UNIV.
13500       EXTERNAL FOF
13600       DIMENSION TH(11), DIFZ(11), SIGNS(11), Y(50)
13700       DIMENSION Q(11), P(11), E(11), PHI(11), TB(11), F(50), R(50)
13800       DIMENSION A(11,11), D(11,11), DELZ(50,11), SER(11)
13900       COMMON /SCR/SER,Q,P,E,PHI,TB,F,R,A,D,DELZ
14000       DO 706 I = NP,10
14100       TH(I+1) =0.0
14200   706 TB(I+1) =0.0
14300    15 CONTINUE
14400    16 DO 19 I = 1,NP
14500       TEMP = DIFZ(I)
14600       IF(TEMP)17,99,18
14700    17 TEMP=-TEMP
14800    18 IF (TEMP.GE.1.0.OR.TH(I).EQ.0.0) GO TO 99
14900    19 CONTINUE
15000       GA=FLAM
15100       NIT = 1
15200       ASSIGN 225 TO IRAN
15300       ASSIGN 265 TO JORD
15400       ASSIGN 180 TO KUWA
15500       IF( EPS1.GE.0.0) GO TO 10
15600     5 EPS1 = 0.0
15700    10 IF( EPS2.GT.0.0) GO TO 30
15800    40 IF( EPS1.GT.0.0) GO TO 50
15900    60 ASSIGN 270 TO IRAN
16000       GO TO 70
16100    50 ASSIGN 265 TO IRAN
16200       GO TO 70
16300    30 IF( EPS1.GT.0.0) GO TO 70
16400    80 ASSIGN 270 TO JORD
16500    70 SSQ =0.
```

```
      CALL FOF(TH,F,NOB,NP)
      DO 90 I =1,NOB
      R(I) = Y(I) - F(I)
      IF (R(I).GT.1.0E10) R(I)=1.0E10
      IF (R(I).LT.-1.0E10) R(I)=-1.0E10
   90 SSQ=SSQ+R(I)*R(I)
      WRITE(6,800) TH,SSQ
  800 FORMAT(40X,"HISTORY OF SEARCH"//7X,"TER FR",6X,"FR 1",5X,"RATE 1",
     *5X,"FR 2",5X,"RATE 2",5X,"FR 3",5X,"RATE 3",5X,"FR 4",5X,"RATE 4",
     *5X,"FR 5",5X,"RATE 5",6X,"SUMSQ"//5X,1P, 11E10.3,3X,E10.3)
C*  ************************
C*      BEGIN ITERATION
C*  ************************
  105 IF(FNU.EQ.0.0) GO TO 99
      GA=GA/FNU
      KCOUNT=0
      DO 130 J=1,NP
      TEMP = TH(J)
      P(J)=DIFZ(J)*TH(J)
      IF (P(J).EQ.0.0) P(J)=1.0E-10
      TH(J)= TH(J)+P(J)
      Q(J)=0.
      CALL FOF(TH,DELZ(1,J),NOB,NP)
      DO 120 I =1,NOB
      DELZ(I,J)=(DELZ(I,J)-F(I))
  120 Q(J)=Q(J)+DELZ(I,J)*R(I)
      IF (P(J).EQ.0.) GO TO 99
      Q(J)= Q(J)/P(J)
C*                              Q=XT*R (STEEPEST DESCENT)
  130 TH(J) = TEMP
      DO 150 I = 1,NP
      DO 151 J=1,I
      SUM =0.
      DO 160 K = 1,NOB
  160 SUM = SUM + DELZ(K,I)*DELZ(K,J)
```

```
20100        TEMP= SUM/(P(I)*P(J))
20200        D(J,I)= TEMP
20300    151 D(I,J)= TEMP
20400 C*                              D=XT*X (MOMENT MATRIX)
20500        E(I) = SQRT(ABS(D(I,I)))
20600        IF (E(I).EQ.0.0) E(I)=1.0E-10
20700    150 CONTINUE
20800        GO TO KUWA, (666,180)
20900 C*                              ITERATION 1 ONLY
21000    180 DO 200 I = 1,NP
21100        DO 200 J = 1,I
21200        SUM = D(I,J)
21300        A(J,I) =SUM
21400    200 A(I,J) =SUM
21500    702 ASSIGN 666 TO KUWA
21600 C*                              END ITERATION 1 ONLY
21700    666 DO 153 I=1,NP
21800        DO 153 J = 1,I
21900        IF (E(I).EQ.0.0.OR.E(J).EQ.0.0) GO TO 99
22000        A(I,J) = D(I,J)/(E(I)*E(J))
22100    153 A(J,I) = A(I,J)
22200 C*                              A = SCALED MOMENT MATRIX
22300        DO 155 I = 1,NP
22400        P(I)=Q(I)/E(I)
22500        PHI(I) =P(I)
22600    155 A(I,I) = A(I,I) + GA
22700        I=1
22800    710 CALL MATINV(A,NP,P,I,DET,NP)
22900 C*                              P/E = CORRECTION FACTOR
23000    701 STEP = 1.0
23100        SUM1=0.
23200        SUM2=0.
23300        SUM3=0.
23400        DO 231 I=1,NP
```

```
23500         SUM1=P(I)*PHI(I)+SUM1
23600         SUM2=P(I)*P(I)+SUM2
23700    231  SUM3=PHI(I)*PHI(I)+SUM3
23800         TEMP= SUM1/SQRT(ABS(SUM2*SUM3))
23900         IF(TEMP.LE.1.0) GO TO 233
24000    232  TEMP= 1.0
24100    233  TEMP= 57.295*ARCOS(TEMP)
24200    170  DO 220 I=1,NP
24300    220  TB(I)=P(I)*STEP/E(I)+TH(I)
24400    704  DO 2401 I=1,NP
24500         IF( SIGNS(I).GT.0.0.AND.TH(I)*TB(I).LE.0.0) GO TO 663
24600   2401  CONTINUE
24700         SUMB=0.
24800         CALL FOF(TB,F,NOB,NP)
24900         DO 230 I=1,NOB
25000         R(I)=Y(I)-F(I)
25100         IF (R(I).GT.1.0E10) R(I)=1.0E10
25200         IF (R(I).LT.-1.0E10) R(I)=-1.0E10
25300    230  SUMB=SUMB+R(I)*R(I)
25400         WRITE(6,801) NIT,TB,SUMB
25500    801  FORMAT(1X,I2,2X,1P, 11E10.3,3X,E10.3)
25600    705  IF(SUMB/SSQ - 1.0.LE. EPS1) GO TO 662
25700    663  IF(TEMP.GT.30.0) GO TO 664
25800    665  STEP=STEP/2
25900         KCOUNT=KCOUNT+1
26000         IF(KCOUNT-36) 170,2700,2700
26100    664  GA=GA*FNU
26200         KCOUNT=KCOUNT+1
26300         IF(KCOUNT-36) 666,2700,2700
26400    662  DO 669 I=1,NP
26500    669  TH(I)=TB(I)
26600         GO TO IRAN, (225,265)
26700    225  DO 240 I=1,NP
26800         IF(ABS (P(I)*STEP/E(I))/(1.0E-20+ABS(TH(I)))-EPS2) 240,240,250
26900    240  CONTINUE
```

```
27000        GO TO 280
27100     250 GO TO JORD, (265,270)
27200     265 IF(ABS((SUMB-SSQ)/SSQ).GT.EPS1) GO TO 270
27300     260 GO TO 280
27400     270 SSQ=SUMB
27500         NIT=NIT+1
27600         IF(NIT-MIT) 105,105,88
27700 C*                                  END ITERATION
27800 C*                                  ************
27900    2700 WRITE (6,85)
28000      85 FORMAT (///10X,  "**** NO CONVERGENCE WITHIN ITERATION ****")
28100     280 SSQ=SUMB
28200         IDF=NOB-NP
28300         I=1
28400     711 CALL MATINV(D,NP,P,I,DET,NP)
28500         DO 7692 I=1,NP
28600    7692 E(I)=SQRT(ABS(D(I,1)))
28700    7058 VAR=SSQ/IDF
28800         SDEV=SQRT(VAR)
28900         DO 391 I=1,NP
29000     391 SER(I)=E(I)*SDEV
29100         IF (NIT.GT.MIT) GO TO 97
29200         WRITE (6,567) SSQ,NIT
29300     567 FORMAT (//5X,"FINAL SUMSQ =",F17.9,10X,"AFTER",I3,2X,"ITERATIONS")
29400      97 WRITE (6,400) VAR,IDF
29500     400 FORMAT (//5X,"VARIANCE =",F20.9,10X,"DEGREES OF FREEDOM = ",I4)
29600         RETURN
29700      88 NN=NIT -1
29800         WRITE (6,568)  SSQ,NN
29900     568 FORMAT (//5X,"FINAL SS =",F20.9,10X,"AFTER",I3,2X,"ITERATION",
30000        * 10X, "(LIMIT REACHED)")
30100         GO TO 280
30200      99 WRITE (6,87) NIT
30300      87 FORMAT (///10X, "**** PARAMETER ERROR OR NO CONVERGENCE ****"
30400        *  ///10X, "**** EXECUTION ENDS AFTER",I4, " ITERATIONS ****")
30500         NN=1
```

```
30600         RETURN
30700         END
30800 C*        ********************
30900 C*
31000 C*
31100         SUBROUTINE MATINV(A,NVAR,B,NB,DETERM,MA)
31200         DIMENSION A(11,11),B(11,1), INDEX(50,2)
31300         EQUIVALENCE (T,SWAP,PIVOT),(K,LI)
31400 C*
31500 C*        INITIALISATION
31600 C*
31700    706  DETERM = 1.0
31800         DO 20 J=1,50
31900     20  INDEX(J,1)=0
32000 C*
32100 C*        SEARCH FOR PIVOT ELEMENT
32200 C*
32300         I=0
32400         IRANK=0
32500     40  AMAX=-1
32600         DO 105 J=1,NVAR
32700         IF (INDEX(J,1)) 105,60,105
32800     60  DO 100 K=1,NVAR
32900         IF (INDEX(K,1)) 100,80,100
33000     80  T=ABS (A(J,K))
33100         IF(T.LE.AMAX) GO TO 100
33200     85  IROW =J
33300         ICOLUM=K
33400         AMAX=T
33500    100  CONTINUE
33600    105  CONTINUE
33700         IF (AMAX) 720,720,110
33800    110  INDEX(ICOLUM,1)=IROW
```

```
C*
C*       INTERCHANGE ROWS TO PUT PIVOT ELEMENT ON DIAGONAL
C*
        IF (IROW.EQ.ICOLUM) GO TO 310
  140 DETERM=-DETERM
        DO 200 L=I,NVAR
        SWAP=A(IROW,L)
        A(IROW,L)=A(ICOLUM,L)
  200 A(ICOLUM,L)=SWAP
        DO 250 L=1,NB
        SWAP=B(IROW,L)
        B(IROW,L)=B(ICOLUM,L)
  250 B(ICOLUM,L)=SWAP
        I=I+1
        INDEX(I,2)=ICOLUM
  310 PIVOT =A(ICOLUM,ICOLUM)
        DETERM=PIVOT*DETERM
        IRANK=IRANK+1
C*
C*       DIVIDE PIVOT ROW BY PIVOT ELEMENT
C*
        A(ICOLUM,ICOLUM)=1.0
        PIVOT=A(ICOLUM,ICOLUM)/PIVOT
        DO 350 L=1,NVAR
  350 A(ICOLUM,L)=A(ICOLUM,L)*PIVOT
        DO 370 L=1,NB
  370 B(ICOLUM,L)=B(ICOLUM,L)*PIVOT
C*
C*       REDUCE NON-PIVOT ROWS
C*
        DO 550 L1=1,NVAR
        IF(L1.EQ.ICOLUM) GO TO 550
  400 T=A(L1,ICOLUM)
        A(L1,ICOLUM)=0.0
        DO 450 L=1,NVAR
```

```
37400    450 A(L1,L)=A(L1,L)-A(ICOLUM,L)*T
37500        DO 500 L=1,NB
37600    500 B(L1,L)=B(L1,L)-B(ICOLUM,L)*T
37700    550 CONTINUE
37800        GO TO 40
37900 C*
38000 C*       INTERCHANGE COLUMS
38100 C*
38200    630 ICOLUM=INDEX(I,2)
38300        IROW=INDEX(ICOLUM,1)
38400        DO 705 K=1,NVAR
38500        SWAP=A(K,IROW)
38600        A(K,IROW)=A(K,ICOLUM)
38700    705 A(K,ICOLUM)=SWAP
38800        I=I-1
38900    720 IF(I) 630,740,740
39000    740 NB=NVAR-IRANK
39100        RETURN
39200         END
39300 C*
39400 C*
39500        SUBROUTINE CRITER(P,CCALC,NRD,NPARAM)
39600        COMMON COT(50),FF(5,50),CCO(50)
39700        DIMENSION CCALC(50),P(11)
39800        DO 2 J=1,NRD
39900      2 CCALC(J)=P(1)
40000        L=0
40100        DO 3 J=3,NPARAM,2
40200        L=L+1
40300        DO 4 K=1,NRD
40400        FF(L,K)= P(J-1)/(1.0000+P(J)*COT(K)))
40500        CCALC(K)=CCALC(K)+FF(L,K)
40600      4 CONTINUE
40700      3 CONTINUE
40800        RETURN
40900         END
```

APPENDIX III

RESTRICTION ENZYME MAPPING

Several of the axolotl recombinant clones were mapped using various restriction enzymes. The basic approach can be demonstrated by the example shown in Fig.AIII.1. Plasmid DNA was isolated as Methods.

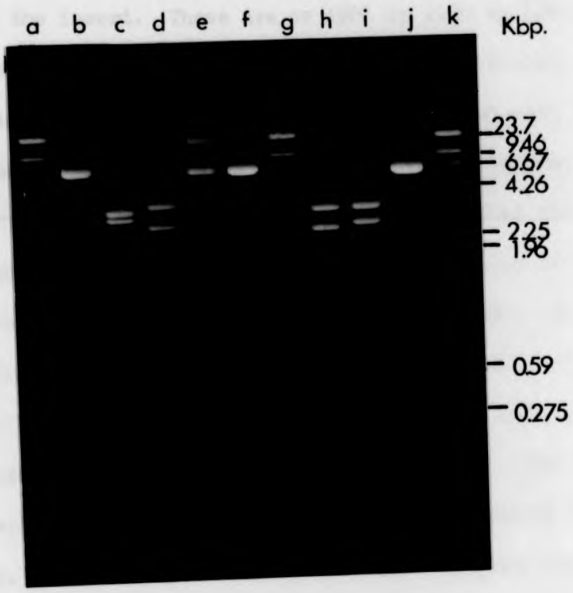
Preliminary digests (e.g. Fig. 5.4) indicated that a number of the clones could not be excised from the BamHI site, indicating damage to one site and resulting in the formation of linear molecules when these plasmids were treated with BamHI. This can be seen, in the case of pAMW1184, in Fig.AIII.1b. A linear fragment of 6.1 kb is produced. The location of the available BamHI site was determined as follows. There is only one site within the plasmid for SalI, 275 bp 3' to the BamHI site in pAT153 (Sutcliffe, 1978) Fig.AIII.2. It was shown that SalI does not cut within the insert (lane f). This was not unexpected as there appears to be few recognition sites for SalI in the genome (see Fig. 5.2 for example). The double digest BamHI + SalI should therefore indicate which BamHI site is available. The 5' BamHI site would release two fragments of 3382 and 2775 bp after double digestion whereas the 3' BamHI site would release two fragments of 5882 and 275 bp after double digestion. The relevant double digest is shown as lane j (Fig. A III.1). It can be clearly seen that a 5.8 kbp and 275 bp fragment are released. This suggests that the BamHI site at the EcoRl proximal end of the insert has been damaged while the EcoRl distal site is still intact. The plasmid was further cleaved with PstI, EcoRl and HindIII (lanes c, d, e, respectively). The HindIII digest (in the case shown the digest has not quite gone to completion) shows the formation of a full length linear molecule indicating only one cleavage site within the whole plasmid and therefore no sites within the insert. The PstI and EcoRl digests both yield two bands of 3225 and 2932 and 3482 and 2675 bp respectively.

LEGEND TO FIG. A III.1

RESTRICTION MAPPING : CLONE 1184

0.5 µg of clone 1184 was digested with a series of restriction enzymes, singly or in pairs. The fragments were separated by electrophoresis on 1% agarose.

a, g, k	λ	-	HindIII
b	1184	-	BamHI
c	1184	-	PstI
d	1184	-	EcoRI
e	1184	-	HindIII
f	1184	-	SalI
h	1184	-	BamHI + PstI
i	1184	-	BamHI + EcoRI
j	1184	-	BamHI + SalI


a b c d e f g h i j k
Kbp.
23.7
9.46
6.67
4.26
2.25
1.98
0.59
0.275

As there is only one site for each enzyme within the plasmid, see Fig. A III.2, this indicates that each enzyme cuts once within the insert.

The distance from the PstI site to the plasmid-insert junction is 1125 bp in the clockwise direction and 2532 bp in the anticlockwise direction. There are therefore two possible locations for the PstI site within the insert. These are at 1807 or 2100 bp (in the clockwise direction) from the EcoRl proximal end of the insert. To discriminate between these two possibilities a double digest with BamHI and PstI was performed. BamHI will only cut within one of the two PstI fragments. The fragment containing the EcoRl proximal section of the insert should remain the same length, however the EcoRl distal section of the insert should be cut free from the plasmid sequences by the BamHI. By this means the two PstI fragments can be oriented and the location of the PstI site within the insert be determined. If the PstI site is at 1807 bp then the fragments produced after the double digest would be 2932, 2532 and 693 bp. If the PstI site is at 2100 bp then the fragments produced would be 3225, 2532 and 400 bp. Lane h (Fig. A III.1) shows the relevant digest. The three visible bands have calculated lengths of 3225, 2532 and 400 bp. Thus the PstI site is 2100 bp from the EcoRl proximal end of the insert.

The EcoRl digest (lane d) also gives two bands of 3482 and 2675 bp. The distances from the EcoRl site in the plasmid to the plasmid insert junctions are 375 bp in the clockwise direction and 3282 bp in the anti-clockwise direction. There is thus only one possible location for the EcoRl site within the insert which would produce the required fragment sizes. This site is 2675 bp in the clockwise direction from the EcoRl site in the plasmid. This was confirmed by a double digest with BamHI + EcoRl. The predicted fragment sizes are 2675, 3282, 200 bp. The relevant digest is shown in lane i. The predicted fragments are clearly visible confirming the position of the EcoRl site.

As well as the enzymes noted above several of the clones were digested

LEGEND TO FIG. A III.2

AID TO RESTRICTION MAPPING; pAT153 FRAGMENT SIZES

The various distances between key restriction sites within the vector pAT153 are shown. These are derived from Sutcliffe (1978). They allow accurate mapping of internal sites within the insert for the enzymes noted

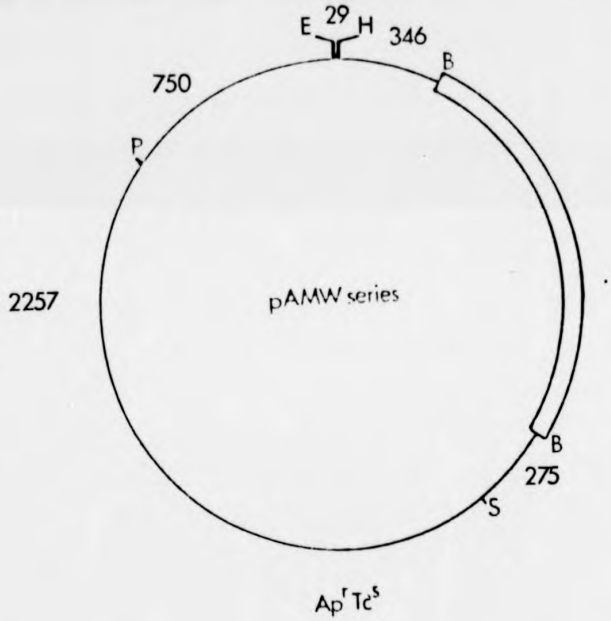
B - BamHI

E - EcoRI

S - SalI

P - PstI

H - HindIII


E
29
H
346
B
750
P
2257
pAMW series
B
275
S
Ap^r Tc^s

with other enzymes including those recognising and cleaving at tetra-nucleotide sites (see for example Fig. 5.6 a - c). Several of the sites could be deduced from these digests, and from partial digest mapping (Smith and Birnstiel, 1978). Most of these sites are still only tentatively assigned and so were not included in Fig. 5.5. These sites are shown in Fig. A III.3.

LEGEND TO FIG. A III.3

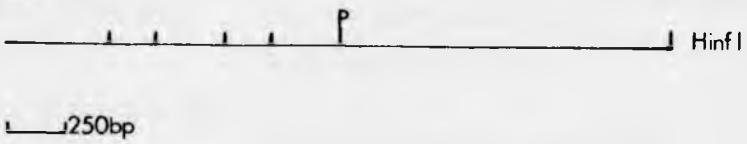
PARTIAL MAPPING DATA

Partial fine restriction mapping data exists for several cloned DNAs. This was not included in the main text as in some cases final map distances have not yet been derived. In some cases restriction sites were located from the analyses shown in Fig. 5.6. In other cases partial end-label mapping (Smith and Birnstiel, 1976) was used.

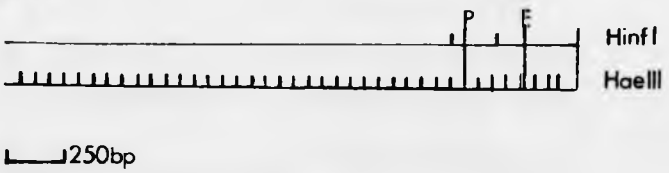
B	BamHl
P	Pstl
S	Sal 1
E	EcoRl
H	HindIII

1131

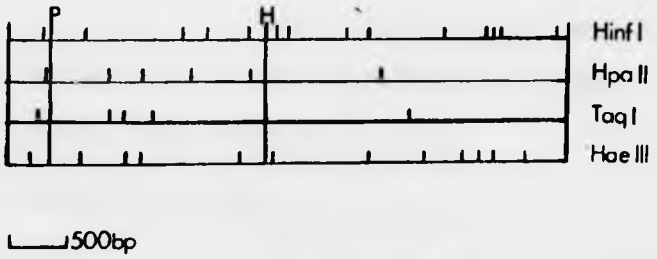
1114



1184



1199



REFERENCES

ALLEN J.R., ROBERTS T.M., LOEBLICH III A.R., KLOTZ L.C.
CELL 6 161 (1975)

ALT F.W., KELLEMS R.E., BERTINO J.R., SCHIMKE R.T.
J.BIOL.CHEM. 253 1357 (1978)

ANDERSON D.M., SCHELLER R.H., POSAKONY J.W., McALLISTER L.B., TRABERT S.G., BEALL C., BRITTEN R.J., DAVIDSON E.H.
J.MOL.BIOL. 145 5 (1981)

ANDERSON D.M., RICHTER J.D., CHAMBERLIN M.E., PRICE D.H., BRITTEN R.J., SMITH L.D., DAVIDSON E.H.
J.MOL.BIOL. 155 281 (1982)

ARTHUR R.R., STRAUS N.A., CAN. J.BIOCHEM. 56 257 (1978)

BACHMANN K. HISTOCHEMIE 22 289 (1970)

BACHMANN K., GOIN O.B., GOIN C.J.
BROOKHAVEN SYMP.BIOL. 23 419 (1972)

BALDARI C.T., AMALDI F. CHROMOSOMA 59 13 (1976)

BALDARI C.T., AMALDI F. CHROMOSOMA 61 359 (1977)

BARSACCHI G., GALL J.G., J.CELL BIOL. 54 580 (1972)

BENDICH A.J., ANDERSON R.S. BIOCHEMISTRY 16 4655 (1977)

BENNETT M.D. in GENOME EVOLUTION p.239
eds. DOVER G.A., FLAVELL R.B.; Pub. ACAD. PRESS 1982

BENNETT M.D., SMITH J.B. PROC. R. SOC. LOND. B. 274 227 (1976)

BERNARDI G. BIOCHIM ET BIOPHYS. ACTA 91 686 (1964)

BIRD A.P. NUCL. ACIDS RES. 8 1499 (1980)

BIRD A.P., TAGGART M.H. NUCL. ACIDS RES. 8 1485 (1980)

BIRD A.P., TAGGART M.H., McLEOD D.
CELL 28 381 (1981)

BIRD A.P., TAGGART M.H., SMITH B.A.
CELL 17 889 (1979)

BISHOP J.O., FREEMAN K.B. COLD SPRING HARBOUR SYMP. QUANT. BIOL. 38 707 (1974)

BOLIVAR F., RODRIGUEZ R.L., GREENE P.J., BETLACH M.C., HEYNECKER H.L., BOYER H.W. GENE 2 95 (1977)

BONNER T., BRENNER D., NEUFELD B., BRITTEN R.J.
J.MOL. BIOL. 81 123 (1973)

BOUCHARD R.A., SWIFT H. CHROMOSOMA 61 317 (1977)

BOYER H.W., ROULLAND-DOUSSOIX D.
J. MOL. BIOL. 41 459 (1969)

BOZZONI I., BECCARI E. BIOCHIM. et BIOPHYS. ACTA 520 245 (1978)

BRIGGS R., CASSENS G. PROC. NATL. ACAD. SCI. 55 1103 (1966)

BRITTEN R.J., DAVIDSON E.H. SCIENCE 165 349 (1969)

BRITTEN R.J., DAVIDSON E.H. PROC. NATL. ACAD. SCI. 73 415 (1976)

BRITTEN R.J., GRAHAM D., NEUFELD B.
in METHODS IN ENZYMOLOGY XXIX E p. 363
eds. GROSSMAN L., MOLDAVE K.; Pub. ACAD. PRESS (1973)

BROTHERS A.J. NATURE 260 112 (1976)

BROWN S.D.M., DOVER G. J. MOL. BIOL. 150 441 (1982)

BRUTLAG D.L., APPELS R., DENNIS E.S., PEACOCK W.J.
J. MOL. BIOL. 112 31 (1977)

BRYKOV V.A., VOLFSON V.G., VOROB'EV V.I.
CHROMOSOMA 74 105 (1979)

BRYKOV V.A., VOLFSON V.G., MOTRENKO E.G., VOROB'EV V.I.
MOL. BIOL. 12 1299 (1978)

CALABRETTA B., ROBBERSON D.L., BARRERA-SALDANA H.A., LAMBRON T.P., SAUNDERS G.F. NATURE 296 219 (1982)

CALLAN H.G. INT. REV. CYTOL. 15 1 (1963)

CALLAN H.G. J. CELL SCI. 1 85 (1966)

CALLAN H.G. PROC. R. SOC. LOND. B 181 19 (1972)

CALLAN H.G., LLOYD L. PHIL TRANS R.SOC. LOND. B 243 135 (1960)

CALLAN H.G., OLD R.W. J. CELL SCI. 41 115 (1980)

CALOS M.P., MILLER J.H. CELL 20 579 (1980)

CAMERON J.R., LOH E.Y., DAVIS R.W.
CELL 16 739 (1979)

CAVALIER-SMITH T. J. CELL SCI. 34 247 (1978)

CAVALIER-SMITH T. NATURE 285 617 (1980)

CHAMBERLIN M.E., GALAU G.A., BRITTEN R.J., DAVIDSON E.H.
NUCL. ACIDS RES. 5 2073 (1978)

CHANG C-T., HAIN T.C., HUTTON J.R., WETMUR J.G.
BIOPOLYMERS 13 1847 (1974)

COEN E.S., THODAY J.M., DOVER G.
NATURE 295 564 (1982)

COLMAN A., EYERS M.J., PRIMROSE S.B., LYONS A.
EUR. J. BIOCHEM. 91 303 (1978)

CONSTANTINI F.D., BRITTEN R.J., DAVIDSON E.H.
NATURE 287 111 (1980)

CRAIN W.R., DAVIDSON E.H., BRITTEN R.J.
CHROMOSOMA 59 1 (1976)b

CRAIN W.R., EDEN F.C., PEARSON W.R., DAVIDSON E.H., BRITTEN R.J.
CHROMOSOMA 56 309 (1976)a

CRAMPTON J.M., DAVIES K.E., KNAPP T.F.
NUCL. ACIDS RES. 9 3821 (1981)

CULLIS C.A. BIOCHIM. ET BIOPHYS. ACTA 652 1 (1981)

CUNY G., SORIANO P., MACAYA G., BERNARDI G.
EUR. J. BIOCHEM. 115 227 (1981)

DAVIDSON E.H. GENE ACTIVITY IN EARLY DEVELOPMENT 2nd Ed.
Pub. ACAD. PRESS (1976)

DAVIDSON E.H., BRITTEN R.J. SCIENCE 204 1052(1979)

DAVIDSON E.H., HOUGH B.R., AMENSON C.S., BRITTEN R.J.
J. MOL. BIOL. 77 1 (1973)

DAVIDSON E.H., HOUGH B.R., KLEIN W.H., BRITTEN R.J.
CELL 4 217 (1974)

DAVIDSON E.H., KLEIN W.H., HOUGH-EVANS B.R., SMITH M.J., GALAU G.A., CRAIN W.R., ANGERER R.C., WOLD B.J., DAVIS M.M., BRITTEN R.J.
in THE ORGANIZATION AND EXPRESSION OF THE EUKARYOTIC GENOME p. 373
eds. BRADBURY E.M., JAVAHERIAN K.; Pub. ACAD. PRESS (1977)

DAVIDSON E.H., POSAKONY J.W. NATURE 297 633 (1982)

DAWID I.B., BROWN, D.D., REEDER R.H.,
J. MOL. BIOL. 51 341 (1970)

DeLAP R., RUSH G.M. PROC. NATL. ACAD. SCI. 75 5855 (1978)

DENHARDT D.T. BIOCHEM. BIOPHYS. RES. COMM. 23 641 (1966)

DIAZ M.O., BARSACCHI-PILONE G., MAHON K.A., GALL J.G.
CELL 24 649 (1981)

DOLECKI G.J., SMITH L.D. DEV. BIOL. 69 217 (1979)

DONS J.J.M., de VRIES O.M.H., WESSELS J.G.H.
BIOCHIM. ET BIOPHYS. ACTA 563 100 (1979)

DONS J.J.M., WESSELS J.G.H. BIOCHIM. ET BIOPHYS. ACTA 607 385 (1980)

DOOLITTLE W.F. in GENOME EVOLUTION p. 3
eds. DOVER G.A., FLAVELL R.B.; Pub. ACAD. PRESS (1982)

DOOLITTLE W.F., SAPIENZA C. NATURE 284 601 (1980)

DOVE W.F., DAVIDSON N. J. MOL. BIOL. 5 467 (1962)a

DOVE W.F., DAVIDSON N. J. MOL. BIOL. 5 479 (1962)b

DOVER G. in CHROMOSOMES TODAY VOL. 6 p. 105
eds. de la CHAPPELLE A., SORSA M.; Pub. ELSEVIER AND N. HOLLAND (1977)

DUGAICZYK A., BOYER H.W., GOODMAN H.M.
J. MOL. BIOL. 96 171 (1975)

EDEN F.C., HENDRICK J.P. BIOCHEMISTRY 17 5838 (1978)

EDEN F.C., MUSTI A.M., SOBIESKI D.A.
J. MOL. BIOL. 148 129 (1981)

EDSTROM J.E., GALL J.G. J. CELL BIOL. 19 279 (1963)

EHRLICH M., WANG R.Y. SCIENCE VOL. 212 p. 1350 (1981)

EICHHORN G.L. NATURE 194 474 (1962)

EFSTRATIADIS A., CRAIN W.R., BRITTEN R.J., DAVIDSON E.H., KAFATOS F.C.
PROC. NATL. ACAD. SCI. 73 2289 (1976)

EPPLEN J.T., DIEDRICH U., WAGENMANN M., SCHMIDTKE J., ENGEL W.
CHROMOSOMA 75 199 (1979)

EPPLEN J.T., LEIPOLDT M., ENGEL W., SCHMIDTKE J.
CHROMOSOMA 69 307 (1978)

FELSENFELD G. in THE MOLECULAR BASIS OF NEOPLASIA p.104
Pub. UNIV. OF TEXAS PRESS, AUSTIN (1962)

FINNEGAN D.J., RUBIN G.M., YOUNG M.W., HOGNESS D.S.
COLD SPRING HARBOUR SYMP. QUANT. BIOL.
42 1053 (1978)

FIRTEL R.A., KINDLE K. CELL 5 401 (1975)

FITZGERALD-HAYES M., CLARKE L., CARBON J.
CELL 29 235 (1982)

FLAVELL A.J., ISH-HOROWICZ D.
NATURE 292 591 (1981)

FLAVELL R.B. in CHROMOSOMES TODAY VOL. 7 p. 42
eds. BENNETT M.D., BOBROW M., HEWITT G.; Pub. ALLEN AND UNWIN (1980)

FLAVELL R.B. in GENOME EVOLUTION p. 301
eds. DOVER G.A., FLAVELL R.B.; Pub. ACAD. PRESS (1981)

FLAVELL R.B., BENNETT M.D., SMITH J.B., SMITH D.B.
BIOCHEM. GENET. 12 257 (1974)

FLAVELL R.B., O'DELL M., SMITH M.,
HEREDITY 42 309 (1979)

FLAVELL R.B., RIMPAU J., SMITH D.B.
CHROMOSOMA 63 205 (1977)

FLAVELL R.B., RIMPAU J., SMITH D.B., O'DELL M., BEDBROOK J.R.
in GENOME ORGANIZATION AND EXPRESSION
IN PLANTS p. 35
ed. LEAVER C.J.; Pub. PLENUM PRESS, NEW YORK (1980)

FLAVELL R.B., SMITH D.B. HEREDITY 37 231 (1976)

FLAVELL R.B., SMITH D.B. NUCL. ACIDS RES. 4 2429 (1977)

FORD P.J., MATHIESON T., ROSBASH M.
DEV. BIOL. 57 419 (1977)

FRY K., SALSER W. CELL 12 1069 (1977)

GALAS D.J., CHANDLER M. PROC. NATL. ACAD. SCI. 78 4858 (1981)

GALAU G.A., CHAMBERLIN M.E., HOUGH B.R., BRITTEN R.J., DAVIDSON E.H.
in MOLECULAR EVOLUTION p. 200
ed. AYALA F.J.; Pub. SINAUER ASSOC. INC. (1976)

GALL J.G., J. MORPHOL. 94 283 (1954)

GILL J.E., MAZRIMAS J.A., BISHOP C.C. Jr.
BIOCHIM. ET BIOPHYS. ACTA 335 835 (1980)

GOLDEN L., SCHAFER U., ROSBASH M.
CELL 22 835 (1980)

GRAHAM D.E., NEUFELD B.R., DAVIDSON E.H., BRITTEN R.J.
CELL 1 127 (1974)

GRAHAM D.E., SCHANKE K. NUCL. ACIDS RES. 8 3875 (1980)

GRAF H. BIOCHIM. ET BIOPHYS. ACTA 564 225 (1979)

GROSSCHEDL R., BIRNSTIEL M.L.
PROC. NATL. ACAD. SCI. 77 7102 (1980)

GRUNSTEIN M., HOGNESS D.S. PROC. NATL. ACAD. SCI. 72 3961 (1975)

GRUNSTEIN M., WALLIS J.J. in METHODS IN ENZYMOLOGY VOL. 68 p. 379
ed. WU R.; Pub. ACAD. PRESS (1979)

HABER J.E., ROGERS D.T. NATURE 296 768 (1982)

HACKET R.W., LIS J.T. PROC. NATL. ACAD. SCI. 78 6196 (1981)

HANHAM A., SMITH M.J. CAN. J. ZOOL. 57 1878 (1979)

HARPOLD M.M., CRAIG S.P. NUCL. ACIDS RES. 12 4425 (1977)

HEILIG R., PERRIN F., GANNON F., MANDEL J.L., CHAMBON P.
CELL 20 625 (1980)

HENNEN S., MIZUNO S., MacGREGOR H.C.
CHROMOSOMA 50 349 (1975)

HERSHFIELD V., BOYER H.W., YANOFSKY C., LOVETT M.A., HELINSKI D.R.
PROC. NATL. ACAD. SCI. 71 3455 (1974)

HILDER V.A., LIVESEY R.N., TURNER P.C., VLAD M.T.
NUCL. ACIDS RES. 9 5737 (1981)

HINEGARDNER R. in MOLECULAR EVOLUTION p. 179
ed. AYALA F.J.; Pub. SINAUER ASSOC. INC. (1976)

HINNEBUSCH A.G., CLARKE V.E., KLOTZ L.C.
BIOCHEMISTRY 17 1521 (1978)

HINNEBUSCH A.G., KLOTZ L.C., IMMERGUT E., LOEBLICH III A.R.
BIOCHEMISTRY 19 1744 (1980)

HOLLIS G.F., HIETER P.A., McBRIDE O.W., SWAN D., LEDER P.
NATURE 296 321 (1982)

HORZ W., ZACHAU H.G., EUR. J. BIOCHEM. 73 383 (1977)

HOUCK C.M., RINEHART F.P., SCHMID C.W.
J. MOL. BIOL. 155 571 (1979)

HOUGH B.R., DAVIDSON E.H. J. MOL. BIOL. 70 491 (1972)

HOURCADE D., DRESSLER P., WOLFSON J.
PROC. NATL. ACAD. SCI. 70 2926 (1973)

HUDSPETH M.E.S., TIMBERLAKE W.E., GOLDBERG
PROC. NATL. ACAD. SCI. 74 4332 (1977)
P

HUMPHRIES P., OLD R.W., COGGINS L.W., McSHANE T., WATSON C., PAUL J.
NUCL. ACIDS RES. 5 905 (1978)

HUTCHINSON J., NARAYAN R.K.J., REES H.
CHROMOSOMA 78 137 (19

HUTTON J.R., WETMUR J.G. J. MOL. BIOL. 77 495 (1973)

KELLS D.I.C., STRAUS N.A. ANAL. BIOCHEM. 80 344 (1977)

KEMP J.D., SUTTON D.W. BIOCHIM. ET BIOPHYS. ACTA. 425 148 (1976)

KEYL H.G. CHROMOSOMA 17 139 (1965)a

KEYL H.G. EXPERIENTIA 21 191 (1965)b

KEZER J., LEON P.E., SESSIONS S.K.
CHROMOSOMA 81 177 (1980)

KIMMEL A.R., FIRTEL R.A. CELL 16 787 (1979)

KINBERG-CALHOUN J., WETMUR J.G.
BIOCHEMISTRY 20 2645 (1981)

KIPER M. ANAL. BIOCHEM. 91 70 (1978)

KIPER M., HERZFELD CHROMOSOMA 65 335 (1978)

KIRK J.T.O. J. MOL. BIOL. 28 171 (1967)

KLEIN W.H., THOMAS T.L., LAI C., SCHELLER R.H., BRITTEN R.J.,
DAVIDSON E.H. CELL 14 889 (1978)

KOHNE D.E., BRITTEN R.J. in PROCEDURES IN NUCLEIC ACIDS RESEARCH p.500
eds. CANTON G.L., DAVIES D.R.; Pub. HARPER AND ROW (1971)

KRUMLAUF R., MARZLUF G.A. BIOCHEMISTRY 18 3705 (1979)

LANGER-SAFER P.R., LEVINE M., WARD D.C.
PROC. NATL. ACAD. SCI. 79 4381 (1982)

LANIA L., BOAST S., FRIED M.
NATURE 295 349 (1982)

LAUER G.D., ROBERTS T.M., KLOTZ L.C.
J. MOL. BIOL. 114 507 (1977)

LEVINE L., GORDON J.A., JENCKS W.P.
BIOCHEMISTRY 2 168 (1963)

LEWIN B. GENE EXPRESSION II 2nd Ed.
Pub. WILEY (1981)

LIS J., NECKAMEYER W., MIRAULT M-E., ARTAVANIS-TSAKONAS S., LALL P.,
MARTIN G., SCHEDL P. DEV. BIOL. 83 291 (1981)

LIVESEY R.N. HONOURS DISSERTATION UNIVERSITY OF WARWICK 1980

LOMEDICO P., ROSENTHAL N., EFSTRATIADIS A., GILBERT W., KOLODNER R.,
TIZARD R. CELL 18 545 (1982)

MACGREGOR H.C. in CHROMATIN AND CHROMATIN STRUCTURE p. 339
eds. LI H.A., ECKHARDT R.; Pub. ACAD. PRESS (1977)

MACGREGOR H.C. CHROMOSOMA 71 57 (1979)

MACGREGOR H.C. HEREDITY 44 3 (1980)

MACGREGOR H.C., ANDREWS C. CHROMOSOMA 63 109 (1977)

MACGREGOR H.C., KEZER J. CHROMOSOMA 33 167 (1971)

MACGREGOR H.C., MIZUNO S., VLAD M.
in CHROMOSOMES TODAY VOL. 5 p.331
eds. PEARSON P.L., LEWIS K.R.; Pub. WILEY AND SONS, NEW YORK (1976)

MALACINSKI G.M. AMERICAN ZOOLOGIST 18 195 (1978)

MALACINSKI G.M., BROTHERS A.J.
SCIENCE 184 1142 (1974)

MANDEL M., MARMUR J. in METHODS IN ENZYMOLOGY XII B p.195
eds. GROSSMAN L., MOLDAVE K.; Pub. ACAD. PRESS (1971)

MANNING J.E., SCHMID C.W., DAVIDSON N.
CELL 4 141 (1975)

MARMUR J., DOTY P. NATURE 183 1427 (1959)

MARMUR J., DOTY P. J. MOL. BIOL. 5 109 (1962)

MARMUR J., ROWND R., SCHILDKRAUT C.L.
in PROGRESS IN NUCLEIC ACIDS RESEARCH 1 231
(1963)

MARMUR J., T'SO P.O.P., BIOCHIM. BI BIOPHYS. ACTA 51 32 (1961)

MARTINSON H.G. BIOCHEMISTRY 12 139 (1973)a

MARTINSON H.G. BIOCHEMISTRY 12 145 (1973)b

McCONNAUGHY B.L., LAIRD C.D., McCARTHY B.J.
BIOCHEMISTRY 8 3289 (1969)

McDONNELL M.W., SIMON M.N., STUDIER F.W.
J. MOL. BIOL. 110 119 (1977)

MIZUNO S., ANDREWS C., MACGREGOR H.C.
CHROMOSOMA 56 1 (1976)

MIZUNO S., MACGREGOR H.C. CHROMOSOMA 48 239 (1974)

MIYAZAWA Y., THOMAS C.A. Jr.,
J. MOL. BIOL. 11 223 (1965)

MOHUN T.J., TILLY R., MOHUN R., SLACK J.M.W.
CELL 22 9 (1980)

MOORE G.P., CONSTANTINI F.D., POSAKONY J.W., DAVIDSON E.H., BRITTEN R.J.
SCIENCE 208 1046 (1980)

MOORE G.P., SCHELLER R.H., DAVIDSON E.H., BRITTEN R.J.
CELL 15 649 (1978)

MORESCALCHI A. in EVOLUTIONARY BIOLOGY VOL. 8 p.339
eds. DOBZHANSKY T., HECHT M.K., STEERE W.C.
Pub. PLENUM PUB. NEW YORK (1980)

MORGAN G.T., MACGREGOR H.C., COLMAN A.
CHROMOSOMA 80 309 (1980)

MOYZIS R.K., BONNET J., LI D.W., T'SO P.O.P.
J. MOL. BIOL. 153 841 (1981)a

MOYZIS R.K., BONNET J., LI D.W., T'SO P.O.P.
J. MOL. BIOL. 153 871 (1981)b

MURRAY M.G., CUELLAR R.E., THOMPSON W.F.,
BIOCHEMISTRY 17 5781 (1978)

MUSTI A.M., SOBIESKI D.A., CHEN B.B., EDEN F.C.
BIOCHEMISTRY 20 2989 (1981)

NISHIOKA Y., LEDER A., LEDER P.
PROC. NATL. ACAD. SCI. 77 2806 (1980)

NORGARD M.V., KEEMS K., MONAHAN J.J.
GENE 3 279 (1978)

NUSSINOV R. NUCL. ACIDS RES. 8 4545 (1980)

NUSSINOV R. J. BIOL. CHEM. 256 8458 (1981)

OLD R.W., PRIMROSE S.B. PRINCIPLES IN GENETIC MANIPULATION 2nd Ed.
Pub. BLACKWELL (1981)

OLD R.W., CALLAN H.G., GROSS K.W.
J. CELL. SCI. 27 57 (1977)

ORGEL L.E., CRICK F.H.C. NATURE 284 604 (1980)

PARDUE M.L., GALL J.G. SCIENCE 168 1356 (1970)

PEACOCK W.J., DENNIS E.S., GERLACH W.L.
in CHROMOSOMES TODAY VOL. 7 p.30
eds. BENNETT M.D., BOBROW M., HEWITT G.M.
Pub. ALLEN AND UNWIN (1980)

PEARSON W.R., WU J-R., BONNER J.
BIOCHEMISTRY 17 51 (1978)

PERLMAN S., PHILLIPS C., BISHOP J.O.,
CELL 8 33 (1976)

POMEROL C. THE CENOZOIC ERA : TERTIARY AND QUARTERNARY
Pub. ELLIS HORWOOD (1982)

POSAKONY J.W., SCHELLER R.H., ANDERSON D.M., BRITTEN R.J., DAVIDSON E.H.
J. MOL. BIOL. 149 41 (1981)

POTTS L. PH.D. THESIS UNIVERSITY OF WARWICK 1981

RAWSON J.R.Y., ECKENRODE V.K., BOERMA C.L., CURTIS S.
BIOCHIM. ET BIOPHYS. ACTA 563 1 (1979)

REES H. BROOKHAVEN SYMP. BIOL. 23 394 (1972)

REES H., JENKINS G., SEAL A.G., HUTCHINSON J.
in GENOME EVOLUTION p. 287
eds. DOVER G.A., FLAVELL R.B.; Pub. ACAD. PRESS (1982)

REES H., JONES R.N. in CHROMOSOME GENETICS VOL. III GENETICS,
PRINCIPLES AND PERSPECTIVES p. 57
eds. LEWIS K.R., JOHN B.; Pub. EDWARD ARNOLD (1977)

REES H., NARAYAN R.K.J. in CHROMOSOMES TODAY VOL. 6 p. 131
eds. de la CHAPPELLE A., SORSA M.; Pub. ELSEVIER & N. HOLLAND (1977)

RICE S.A., DOTY P. J. AM. CHEM. SOC. 79 3937 (1957)

RIGBY P.W.J., DIECKMANN M., RHODES C., BERG P.
J. MOL. BIOL. 113 237 (1977)

RIMPAU J., SMITH D.B., FLAVELL R.B.
J. MOL. BIOL. 123 327 (1978)

RIMPAU J., SMITH D.B., FLAVELL R.B.,
HEREDITY 44 131 (1980)

RINEHART F.P., RITCH T.G., DEININGER P.L., SCHMID C.W.
BIOCHEMISTRY 20 3003 (1981)

ROBERTS R. NUCL. ACIDS RES. 11 1135 (1983)

ROSBASH M., FORD P.J., BISHOP J.O.
PROC. NATL. ACAD. SCI. 71 3746 (1974)

RYFFEL G.U., MUELLENER D.B., WYLER T., WAHLIE W., WEBER R.
NATURE 291 429 (1981)

SALSER W., ISAACSON J.S. PROGRESS IN NUCLEIC ACIDS RES. 19 205 (1976)

SAMOLS D., SWIFT H. CHROMOSOMA 75 129 (1979)

SCHACHAT F.H., HOGNESS D.S. in COLD SPRING HARBOUR SYMP. QUANT. BIOL. VOL. 38 p.371 (1973)

SCHELLER R.H., ANDERSON D.M., POSAKONY J.W., McALLISTER L.B., BRITTEN R.J., DAVIDSON E.H. J. MOL. BIOL. 175 15 (1981)

SCHELLER R.H., CONSTANTINI F.D., KOZLOWSKI M.R., BRITTEN R.J. DAVIDSON E.H. CELL 15 189 (1978)

SCHELLER K., MASHENK P., HOPPE-SEYLERS Z.
PHYSIOL. CHEM. 359 1413 (1978)

SCHIMKE R.T. in CHROMOSOMES TODAY VOL. 7 p.24
eds. BENNETT M.D., BOBROW M., HEWITT G.M.; Pub. ALLEN & UNWIN (1980)

SCHMID C.W., DEININGER P.L. CELL 6 345 (1975)

SCHMIDTKE J., EPPLEN J.T., ENGEL W.
COMP. BIOCHEM. PHYSIO. 63B 445 (1979)

SCHMIDTKE J., SCHMITT E., MATZKE E., ENGEL E.
CHROMOSOMA 75 185 (1979)

SESSIONS S.K. CHROMOSOMA 84 599 (1982)

SHAPIRO J.A. PROC. NATL. ACAD. SCI. 76 1933 (1979)

SINCLAIR J.H., CARROLL C.R., HUMPHREY R.R.
J. CELL. SCI. 15 239 (1974)

SMITH C.A., VINOGRAD J. J. MOL. BIOL. 69 163 (1972)

SMITH G.P. SCIENCE 191 528 (1976)

SMITH D.B., FLAVELL R.B. BIOCHEM. GENET. 12 243 (1974)

SMITH D.B., RIMPAU J., FLAVELL R.B.
NUCL. ACIDS RES. 3 2811 (1976)

SMITH H.O., BIRNSTIEL M.L. NUCL. ACIDS RES. 3 2387 (1976)

SMITH M.J., BRITTEN R.J., DAVIDSON E.H.
PROC. NATL. ACAD. SCI. 71 4805 (1975)

SOBIESKI D.A, EDEN F.C. NUCL. ACIDS RES. 9 6001 (1981)

SOMMERVILLE J. in INT. REV. BIOCHEM. VOL. 15. 79.
ed. PAUL J.; Pub. UNIV. PARK PRESS BALTIMORE (1977)

SOMMERVILLE J., MALCOLM D.B. CHROMOSOMA 55 183 (1976)

SORIANO P., MACAYA G., BERNARDI G.
EUR. J. BIOCHEM. 115 235 (1981)

SOUTHERN E. NATURE 232 82 (1971)

SOUTHERN E. J. MOL. BIOL. 98 503 (1975)

SPARROW A.H., PRICE H.J., UNDERBRINK A.G. BROOKHAVEN SYMP. BIOL. 23 451 (1972)

SPOHR G., REITH W., SURES I. J. MOL. BIOL. 151 573 (1981)

SPRADLING A.C., RUBIN G.M. ANN. REV. GENET. 15 219 (1981)

STANFIELD S.W., LENGYEL J.A. PROC. NATL. ACAD. SCI. 76 6142 (1979)

STEPHENSON Jr. E.C. PH.D. THESIS YALE UNIVERSITY NEWHAVEN CONNECTICUT (1981)

STEPHENSON E.C., ERBA H.C., GALL J.G. CELL 24 639 (1981)

STRAUS N.A. PROC. NATL. ACAD. SCI. 68 799 (1971)

STRINGER J.R. NATURE 296 363 (1982)

SWIFT H. PROC. NATL. ACAD. SCI. 36 643 (1950)

SISSOËFF I., GRISVARD J., GUILLE E. PROG. BIOPHYS. MOL. BIOL. 31 165 (1976)

SUBIRIANA J.A., DOTY P. BIOPOLYMERS 4 171 (1966)

SUTCLIFFE J.G. NUCL. ACIDS RES. 5 2721 (1978)

SZYBALSKI W., SZYBALSKI E.F. in PROCEDURES IN NUCLEIC ACID RESEARCH VOL. 2 p.338
eds. CANTONI G.L., DAVIES B.R.; Pub. HARPER AND ROW (1971)

TIHEN J.A. BULL. FLA. ST. MUS. BIOL. 3 1 (1958)

TIMBERLAKE W.E. SCIENCE 202 973 (1978)

THIERY J-P., MACAYA G., BERNARDI G. J. MOL. BIOL. 108 219 (1976)

THOMAS C.A. ANN. REV. GENET. 5 237 (1971)

THOMPSON W.F. PLANT PHYSIOL. 57 617 (1976)

THOMPSON W.F., MURRAY M.G., CUELLAR R.E. in GENOME ORGANIZATION AND EXPRESSION IN PLANTS p.1.
ed. LEAVER C.J.; Pub. PLENUM N.Y. (1980)

THOMPSON W.F., PRIESLER R.S. J. MOL. EVOL. 17 85 (1981)

T'SO P.O.P., HELMKAMP G.K., SANDER C. PROC. NATL. ACAD. SCI. 48 686 (1962)a.

O'SO P.O.P., HELMKAMP G.K., SANDER C.
BIOCHIM. ET BIOPHYS. ACTA 55 564 (1962)b.

TUNIS M.J.P., HEARST J.E. BIOPOLYMERS 6 1325 (1973)

TURNER P.C., WOODLAND H.R. NUCL. ACIDS RES. 11 971 (1983)

TWIGG A.J., SHERRATT D. NATURE 283 216 (1980)

ULLRICH R.C., KOHORN B.L., SPECHT C.A.
CHROMOSOMA 81 371 (1980)

VAN ARSDELL S.W., DENISON R.A., BERNSTEIN L.B., WEINER A.M., MANSER T., GESTELAND R.F. CELL 26 11 (1981)

VANIN E.F., GOLDBERG G.I., TUCKER P.W., SMITHIES O.
NATURE 286 222 (1980)

VANYUSHIN B.F., TKACHEVA S.G., BELOZERSKY A.N.
NATURE 225 948 (1970)

VARLEY J.M., MACGREGOR H.C., ERBA H.
NATURE 283 686 (1980)a.

VARLEY J.M., MACGREGOR H.C., NARDI I., ANDREWS C., ERBA H.C.
CHROMOSOMA 80 289 (1980)b.

VOGT V.M. EUR. J. BIOCHEM. 33 192 (1973)

WAALWIJK C., FLAVELL R.A. NUCL. ACIDS RES. 5 3231 (1978)

WAGNER I., CAPESIUS I. BIOCHIM. ET BIOPHYS. ACTA 654 52 (1981)

WALBOT V., DURE III L.S. J. MOL. BIOL. 101 503 (1976)

WALKER P.M.B. PROG. BIOPHYS. MOL. BIOL. VOL. 23 145 (1971)

WELLS R., ROYER H-D., HOLLENBERG C.
MOL. GEN. GENET. 147 45 (1976)

WENSINK P.C. COLD SPRING HARBOUR SYMP. QUANT. BIOL. 42 1033 (1977)

WENSINK P.C., TAVATA S., PACHL C.
CELL 18 1231 (1979)

WETMUR J.G. BIOPOLYMERS 10 601 (1971)

WETMUR J.G., DAVIDSON N. J. MOL. BIOL. 31 349 (1968)

WHITNEY P.A., HALL B.D. FED. PROC. 33 1282 (1974)

WHITTAKER P.A., McLACHLAN A., HARDMAN N.
NUCL. ACIDS RES. 9 801 (1981)

WILSON D.A., THOMAS Jr. C.A. J. MOL. BIOL. 84 115 (1974)

WOBUS U. EUR. J. BIOCHEM. 59 287 (1975)

WYATT G.R. BIOCHEM. J. 48 584 (1951)